

A.D. Aleksandrov, A.N. Kolmogorov,
M.A. Laurentiev y otros

La matemática: 1. su contenido, métodos y significado

Alianza Universidad



LA MATEMATICA: SU CONTENIDO, METODOS Y SIGNIFICADO ofrece una panorámica difícilmente mejorable en ese vasto campo temático. Publicada originalmente en la Unión Soviética, su traducción al inglés —bajo el patrocinio de la «American Mathematical Society»— ha sido saludada con grandes elogios por los críticos especializados: «una principalísima contribución a la cultura científica de nuestro siglo» (Harry Schwartz en «The New York Times Book Review»); «la mejor respuesta dada en forma de libro para la enseñanza de las matemáticas tanto en nivel elemental como en el avanzado» (Morris Kline); «una formidable y fabulosa obra maestra» («Social Science»). El tratado se mueve en dos planos: el histórico, que examina el desarrollo de la disciplina desde los orígenes hasta nuestros días, subraya la relación directísima, eclipsada en ocasiones por su alto nivel de abstracción, que las matemáticas guardan con el mundo material; y el expositivo, que logra una organización a gran escala de la matemática moderna. Los tres volúmenes (AU 68, 69 y 70) se subdividen en tres partes y diecinueve capítulos, que se ocupan de las principales ramas: análisis, geometría analítica, teoría de las ecuaciones algebraicas, ecuaciones diferenciales ordinarias, curvas y superficies, cálculo de variaciones, números primos, teoría de probabilidades, aproximación de funciones, álgebra lineal, geometrías no euclidianas, topología, análisis funcional, etcétera. P.S. ALEKSANDROV, I.M. GELFAND, A.S. KOLMOGOROV, M.A. LAURENTIEV y el resto de los diecisiete autores que han colaborado en la empresa —«los más destacados nombres de las matemáticas rusas», ha señalado «Science Fortnightly»— discutieron el plan general de la obra y las sucesivas versiones de cada capítulo, solicitando en determinados casos la opinión y colaboración de otros matemáticos soviéticos. La obra ofrece también grandes virtualidades didácticas; como ha señalado Roger S. Pinkham, «su elevada calidad expositiva y su insistencia en los conceptos más que en las demostraciones contribuirán decisivamente a elevar el nivel matemático de todos los que la estudien».

Alianza Editorial

ISBN 84-206-2068-8



Cubierta: Daniel Gil

9 788420 620688

La matemática:
su contenido,
métodos y
significado
I.

Alianza Universidad

A. D. Aleksandrov
A. N. Kolmogorov,
M. A. Laurentiev
y otros

La matemática:
su contenido,
métodos y
significado

I.

Versión española de
Manuel López Rodríguez

Alianza
Editorial

Título de la versión inglesa:
Mathematics: Its Content, Methods, and Meaning

Primera edición en «Alianza Universidad»: 1973
Décima reimpresión en «Alianza Universidad»: 2003

Reservados todos los derechos. De conformidad con lo dispuesto en el art. 534 bis del Código Penal vigente, podrán ser castigados con penas de multa y privación de libertad quienes reprodujeren o plagiaren, en todo o en parte, una obra literaria, artística o científica fijada en cualquier tipo de soporte sin la preceptiva autorización.

© The Massachussets Institute of Technology
© Ed. cast.: Alianza Editorial, S. A., Madrid, 1973, 1976, 1979, 1980, 1981, 1982, 1985, 1988, 1991, 1994, 2003
Calle Juan Ignacio Luca de Tena, 15; 28027 Madrid; teléf. 91 393 88 88
www.alianzaeditorial.es
ISBN: 84-206-2993-6 (Obra completa)
ISBN: 84-206-2068-8 (Tomo I)
Depósito legal: M. 5.456-2003
Impreso en Closas-Orcoyen, S. L. Polígono Igarsa
Paracuellos de Jarama (Madrid)
Printed in Spain

INDICE

PROLOGO A LA EDICION RUSA.....	11
--------------------------------	----

Parte 1

1. VISION GENERAL DE LA MATEMATICA

A. D. Aleksandrov

§1. Rasgos característicos de la matemática.....	17
§2. Aritmética.....	24
§3. Geometría.....	38
§4. Aritmética y geometría.....	43
§5. La era de la matemática elemental.....	56
§6. La matemática de las magnitudes variables.....	65
§7. La matemática contemporánea.....	79

2. ANALISIS

M. A. Laurentiev y S. M. Nikolski

§1. Introducción.....	91
§2. Función.....	100
§3. Límites.....	108
§4. Funciones continuas.....	117
§5. Derivada.....	121
§6. Reglas para la derivación.....	132

§7. Máximos y mínimos; estudio de la gráfica de una función.....	140
§8. Incremento y diferencial de una función	150
§9. Fórmula de Taylor	157
§10. Integral	163
§11. Integrales indefinidas; técnica de integración	173
§12. Funciones de varias variables.	179
§13. Generalizaciones del concepto de integral.....	196
§14. Series.....	206

Parte 2

3. GEOMETRIA ANALITICA

B. N. Delone

§1. Introducción.....	225
§2. Los dos conceptos fundamentales de Descartes.....	226
§3. Problemas elementales.....	229
§4. Discusión de las curvas representadas por ecuaciones de primer y segundo grado	231
§5. Método de Descartes para la resolución de las ecuaciones algebraicas de tercer y cuarto grado.....	234
§6. Teoría general de los diámetros de Newton.....	237
§7. Elipse, hipérbola y parábola	240
§8. Reducción de la ecuación general de segundo grado a forma canónica.....	353
§9. Representación de las fuerzas, velocidades y aceleraciones por ternas de números: teoría de vectores.....	260
§10. Geometría analítica del espacio; ecuaciones de una superficie y de una curva en el espacio.....	266
§11. Transformaciones afín y ortogonal.....	276
§12. Teoría de los invariantes.....	288
§13. Geometría proyectiva	293
§14. Transformaciones de Lorentz.....	301
Conclusión.....	310

4. ALGEBRA: TEORIA DE LAS ECUACIONES ALGEBRAICAS

B. N. Delone

§1. Introducción.....	315
§2. Solución algebraica de una ecuación.....	320
§3. El teorema fundamental del álgebra.....	338
§4. Estudio de la distribución de las raíces de un polinomio sobre el plano complejo.....	352
§5. Cálculo aproximado de raíces.....	364

5. ECUACIONES DIFERENCIALES ORDINARIAS

I. G. Petrovski

§1. Introducción.....	373
§2. Ecuaciones diferenciales lineales con coeficientes constantes.....	386
§3. Observaciones generales sobre la formación y solución de las ecuaciones diferenciales.....	395
§4. Interpretación geométrica del problema de integrar ecuaciones diferenciales; generalización del problema.....	398
§5. Existencia y unicidad de la solución de una ecuación diferencial; solución aproximada de ecuaciones.....	401
§6. Puntos singulares.....	410
§7. Teoría cualitativa de las ecuaciones diferenciales ordinarias....	415

PROLOGO A LA EDICION RUSA

Las matemáticas, surgidas en la Antigüedad por necesidades de la vida cotidiana, se han convertido en un inmenso sistema de variadas y extensas disciplinas. Como las demás ciencias, reflejan las leyes del mundo que nos rodea y sirven de potente instrumento para el conocimiento y dominio de la naturaleza. Pero el alto nivel de abstracción que caracteriza a las matemáticas trae consigo el que todas sus ramas sean relativamente inaccesibles a los no especialistas. Esta cualidad abstracta de las matemáticas dio lugar, ya en la Antigüedad, a nociones idealistas sobre su independencia respecto del mundo material.

En la preparación del presente volumen, la finalidad perseguida por los autores ha sido la de familiarizar a un círculo suficientemente amplio de la intelligentsia soviética con las diversas disciplinas matemáticas, su contenido y métodos, los fundamentos sobre los que descansan y los caminos a lo largo de los cuales se han desarrollado.

El bagaje mínimo de conocimientos matemáticos que precisa el lector de este libro son las matemáticas a nivel de enseñanza media, aunque los volúmenes difieren en lo que toca a la accesibilidad del material tratado. El lector que desee iniciarse en los elementos de las matemáticas superiores puede leer con provecho algunos de los

primeros capítulos; sin embargo, para una completa comprensión de las partes subsiguientes le será necesario recurrir a los correspondientes libros de texto. El libro en su totalidad podrá ser asimilado de manera fundamental solamente por aquellos lectores que ya tengan algún conocimiento de las aplicaciones del análisis matemático, es decir, del cálculo diferencial y del cálculo integral. Para tales lectores —profesores de matemáticas, ingeniería y ciencias naturales— será particularmente importante la lectura de aquellos capítulos en los que se introducen las ramas más recientes de las matemáticas.

Naturalmente, dentro de los límites de un solo libro no ha sido posible agotar toda la riqueza ni siquiera de los resultados más fundamentales de la investigación matemática; una cierta libertad en la elección del material ha sido inevitable. Pese a ello, en líneas generales el presente libro dará una idea de la situación actual de las matemáticas, sus orígenes y su probable desarrollo futuro. Por esta razón, la obra está pensada también hasta cierto punto para personas ya familiarizadas con la mayoría de los temas contenidos en ella, lo cual quizá contribuya a eliminar esa estrechez de miras que se descubre ocasionalmente en algunos de nuestros jóvenes matemáticos.

Los distintos capítulos del libro están escritos por diversos autores cuyos nombres se dan en el índice; no obstante, el libro en su conjunto ha sido escrito en colaboración. El plan general, la elección del material, las sucesivas versiones de capítulos concretos, todo esto fue sometido a discusión general, introduciendo mejoras aquí y allá tras un animado cambio de opiniones. Matemáticos de distintas ciudades de la Unión Soviética tuvieron la oportunidad, mediante discusiones organizadas, de hacer numerosas y valiosas observaciones concernientes a la versión original del texto. Sus opiniones y sugerencias han sido tenidas en cuenta por los autores.

Los autores de algunos capítulos también tomaron parte directa en la redacción de la versión final de otros capítulos: la introducción del capítulo 2 fue escrita esencialmente por B. N. Delone, mientras que D. K. Fadiev desempeñó un activo papel en la preparación de los capítulos 4 y 18. Parte del trabajo fue también realizado por personas distintas de los autores de los diferentes capítulos: la sección 6 del capítulo 6 fue redactada por O. A. Ladyzenskaia; la sección 5 del capítulo 10, por A. G. Postnikov; sobre el texto del capítulo 5 trabajó O. A. Oleinik, y sobre el del capítulo 11, Ju. V. Prohorov.

Ciertas secciones de los capítulos 1, 2, 7, 16 fueron escritas por V. A. Zalgaller. La edición del texto final fue realizada por V. A. Zalgaller y V. S. Videnski con la colaboración de T. V. Rogozkinaia y A. P. Leonovaia.

La mayor parte de las ilustraciones fueron preparadas por E. P. Senikin.

Moscú
1956

COMITÉ DE EDITORES

Parte I

Capítulo I

VISION GENERAL DE LA MATEMATICA

Una adecuada presentación de cualquier ciencia no puede consistir sólo en información detallada, aunque sea extensa; debe también dar una visión propia de la naturaleza esencial de la ciencia en conjunto. El objeto del presente capítulo es dar un cuadro general de la naturaleza esencial de la matemática. Para ello no hay gran necesidad de entrar en detalles de teorías matemáticas recientes, puesto que la matemática elemental y la historia de la ciencia ya proporcionan base suficiente para obtener conclusiones generales.

§ 1. *Rasgos característicos de la matemática*

1. *Abstracciones, demostraciones, aplicaciones*

Incluso con un conocimiento superficial de la matemática, es fácil reconocer ciertos rasgos característicos: su abstracción, su precisión, su rigor lógico, el irrefutable carácter de sus conclusiones y, finalmente, el campo excepcionalmente amplio de sus aplicaciones.

Es fácil reconocer el carácter abstracto de la matemática. Operamos con números abstractos sin preocuparnos de cómo relacionarlos

en cada caso a objetos concretos. En la escuela se estudia la tabla abstracta de multiplicar, esto es, una tabla para multiplicar un número abstracto por otro, no un número de muchachos por un número de manzanas o un número de manzanas por el precio de una manzana.

De modo similar, en geometría consideramos, por ejemplo, líneas rectas y no hilos gruesos, llegándose al concepto de línea geométrica por abstracción de todas las propiedades excepto la extensión en una dirección. En general, el concepto de figura geométrica es el resultado de la abstracción de todas las propiedades de un objeto exceptuadas su forma espacial y dimensiones.

Abstracciones de esta clase son características en toda la matemática. Los conceptos de número y de figura geométrica son sólo dos de sus primeros y más elementales ejemplos, seguidos luego por muchos otros, demasiado numerosos para describirlos, y que tienen que ver con abstracciones tales como números complejos, funciones, integrales, diferenciales, funcionales, espacios n -dimensionales e incluso espacios de infinitas dimensiones, etc. Estas abstracciones, apoyadas unas en otras, han alcanzado tal grado de generalización que pierden aparentemente toda conexión con la vida diaria, y el hombre medio no entiende nada de ellas salvo el simple hecho de que «todo esto es incomprensible».

La realidad, naturalmente, no es esa en absoluto. Aunque el concepto de espacio n -dimensional es, sin duda, extremadamente abstracto, todavía tiene un contenido completamente real, que no es muy difícil de entender. En este libro nuestro propósito será el de dar a conocer y aclarar el contenido concreto de conceptos abstractos tales como los ya mencionados, de modo que el lector pueda convencerse por sí mismo de que todos ellos están relacionados con la vida real, tanto en su origen como en sus aplicaciones.

Pero la abstracción no es algo exclusivo de la matemática; es característica de toda ciencia, incluso de toda actividad mental en general. Consecuentemente, la abstracción de los conceptos matemáticos no proporciona por sí misma una descripción del carácter peculiar de la matemática.

Las abstracciones de la matemática se distinguen por tres rasgos. En primer lugar, tratan fundamentalmente de las relaciones cuantitativas y formas espaciales, abstrayéndolas de todas las demás propiedades de los objetos. En segundo lugar, aparecen en una sucesión de grados de abstracción creciente, llegando mucho más lejos en esta dirección que la abstracción en las demás ciencias. Más tarde ilus-

traremos estas dos cualidades en detalle, empleando como ejemplos las nociones fundamentales de número y figura. Finalmente, y esto es obvio, la matemática como tal se mueve casi por completo en el campo de los conceptos abstractos y sus interrelaciones. Mientras el científico de la naturaleza experimenta constantemente para demostrar sus aseveraciones, el matemático emplea sólo razonamientos y cálculos.

Es cierto que los matemáticos también hacen constante uso —como ayuda en el descubrimiento de teoremas y métodos— de modelos y analogías físicas, y que recurren con frecuencia a ejemplos bien concretos. Estos constituyen la fuente real de la teoría y un medio de descubrir teoremas; pero ningún teorema pertenece definitivamente a la matemática hasta que no ha sido rigurosamente demostrado por un razonamiento lógico. Si un geómetra diese cuenta de un teorema recientemente descubierto mediante simples modelos y se limitara a tal demostración, ningún matemático admitiría que el teorema había sido probado. La necesidad de demostrar los teoremas es ya normal en la geometría del bachillerato y se extiende a toda la matemática. Podríamos medir los ángulos de la base de miles de triángulos isósceles con extrema precisión, pero ese procedimiento nunca nos daría una demostración matemática del teorema que dice que esos dos ángulos son iguales. La matemática pide que este resultado se deduzca de los conceptos fundamentales de la geometría, conceptos que, teniendo en cuenta el hecho de que la geometría de nuestros días está desarrollada sobre una base rigurosa, se hallan formulados con toda precisión en los axiomas. Y así es en todos los casos. Demostrar un teorema significa que el matemático lo deduzca, mediante un razonamiento lógico, a partir de propiedades fundamentales de los conceptos que aparecen en ese teorema. De este modo, no sólo los conceptos, sino también los métodos de la matemática, son abstractos y teóricos.

Los resultados de la matemática se distinguen por su alto grado de rigor lógico, y los razonamientos matemáticos se desarrollan con una minuciosidad tal que lo hagan incontestable y convincente para todo el que lo entienda. La minuciosidad y fuerza de las demostraciones matemáticas son ya bien conocidas en los cursos de bachillerato superior. Las verdades matemáticas son de hecho el prototipo de lo completamente incontestable. Por algo se dice: «tan claro como que dos y dos son cuatro». Aquí la relación «dos y dos son cuatro» se emplea como paradigma de lo irrefutable e incontestable.

Pero el rigor de la matemática no es absoluto; está en proceso

de continuo desarrollo; los principios de la matemática no se han congelado de una vez para siempre sino que tienen su propia vida y pueden incluso ser objeto de discusiones científicas.

En último término la vitalidad de la matemática se debe al hecho de que, a pesar de su abstracción, sus conceptos y resultados tienen su origen, como veremos, en el mundo real y encuentran muchas y diversas aplicaciones en otras ciencias, en ingeniería y en todos los aspectos prácticos de la vida diaria; reconocer esto es el requisito previo más importante para entender la matemática.

La excepcional amplitud de sus aplicaciones es otro rasgo característico de la matemática.

En primer lugar hacemos constante uso, en la industria y en la vida social y privada, de los más variados conceptos y resultados de la matemática sin pensar en ello; por ejemplo, empleamos la aritmética para calcular nuestros gastos o la geometría para calcular la superficie de un apartamento. Naturalmente, las reglas a emplear son muy **sencillas, pero deberíamos recordar que en algún período de la antigüedad representaron los logros matemáticos más avanzados de la época.**

Segundo, la tecnología moderna sería imposible sin la matemática. No hay probablemente un sólo proceso técnico que pueda realizarse sin cálculos más o menos complicados; y la matemática juega un papel muy importante en el desarrollo de nuevas ramas de la tecnología.

Finalmente, es cierto que toda ciencia, en mayor o menor grado, hace un uso esencial de la matemática. Las «ciencias exactas», mecánica, astronomía, física y una gran parte de la química, expresan sus leyes, como todo estudiante sabe, por medio de fórmulas, y utilizan ampliamente el aparato matemático en el desarrollo de sus teorías. El progreso de estas ciencias habría sido completamente imposible sin la matemática. Por esta razón, las necesidades de la mecánica, astronomía y física han ejercido siempre una directa y decisiva influencia en el desarrollo de la matemática.

En otras ciencias la matemática tiene un papel menor, pero también encuentra importantes aplicaciones. Naturalmente, en el estudio de fenómenos tan complicados como los que aparecen en biología y sociología, los métodos matemáticos no pueden desempeñar el mismo papel que, por ejemplo, en la física. En todos los casos, pero especialmente allí donde los fenómenos son más complicados, debemos tener en cuenta que si no queremos perder el tiempo manejando fórmulas desprovistas de significado, la aplicación de la matemática

es útil sólo si se aplica a fenómenos concretos que ya han sido objeto de una profunda teoría. De un modo u otro, la matemática se aplica en casi todas las ciencias, desde la mecánica hasta la economía política.

Recordemos algunas aplicaciones particularmente brillantes de la matemática en las ciencias exactas y en la tecnología.

El planeta Neptuno, uno de los más distantes del sistema solar, fue descubierto en el año 1846 mediante cálculos matemáticos. Analizando ciertas irregularidades en el movimiento de Urano, los astrónomos Adams y Leverrier llegaron a la conclusión de que estas irregularidades eran producidas por la atracción gravitatoria de otro planeta. Leverrier calculó, basándose en las leyes de la mecánica, el lugar exacto donde debía estar el planeta; y un observador a quien comunicó sus resultados lo localizó con su telescopio en la posición indicada por Leverrier. Este descubrimiento fue un triunfo no sólo para la mecánica y la astronomía (y en particular para el sistema de Copérnico), sino también para la potencia del cálculo matemático.

Otro ejemplo, no menos impresionante, fue el descubrimiento de las ondas electromagnéticas. Generalizando las leyes de los fenómenos electromagnéticos establecidos experimentalmente, el físico inglés Maxwell logró expresarlas en forma de ecuaciones. A partir de estas ecuaciones dedujo, por métodos puramente matemáticos, que las ondas electromagnéticas podían existir y que debían propagarse con la velocidad de la luz. A partir de este resultado propuso la teoría electromagnética de la luz, que fue más tarde desarrollada y profundizada en todas direcciones. Además, los resultados de Maxwell llevaron a la búsqueda de ondas electromagnéticas de origen puramente eléctrico, obteniéndose, por ejemplo, de una carga oscilante. Estas ondas fueron observadas experimentalmente por Hertz. Poco después, A. S. Popov, al descubrir el modo de excitar, transmitir y recibir oscilaciones electromagnéticas las hizo útiles para un gran campo de aplicaciones y de ese modo puso los cimientos de toda la radiotécnica. En el descubrimiento de la radio, ahora posesión común de todos, tuvieron un papel muy importante los resultados de una deducción puramente matemática.

Así, partiendo de la observación —como por ejemplo de la desviación de una aguja magnética por una corriente eléctrica—, la ciencia procede por generalización a una teoría de los fenómenos, a una formulación de las leyes y a expresiones matemáticas de ellas. De estas leyes vienen nuevas deducciones y, finalmente, la teoría es llevada a la práctica, que a su vez proporciona nuevos y poderosos impulsos al desarrollo de la teoría.

Es particularmente notable que incluso las construcciones más abstractas de la matemática, aquellas que surgen dentro de la misma ciencia sin motivación inmediata de las ciencias naturales o de la tecnología, tienen sin embargo fructíferas aplicaciones. Por ejemplo, los números imaginarios vieron por primera vez la luz en álgebra, y durante largo tiempo su significado en el mundo real permaneció desconocido, circunstancia indicada por su propio nombre. Pero cuando alrededor de 1800 se les dio una interpretación geométrica (ver cap. 4, § 2), los números imaginarios quedaron firmemente afincados en la matemática, dando lugar a la extensa teoría de funciones de una variable compleja, es decir, de una variable de la forma $x + y\sqrt{-1}$. Esta teoría de funciones «imaginarias» de una variable «imaginaria» demostró que, lejos de ser imaginaria, era un medio muy práctico de resolver problemas tecnológicos. Así, los resultados fundamentales de N. E. Jukovski referentes a la fuerza ascensional del ala de un aeroplano en vuelo se demostraron por medio de esta teoría. La misma teoría es útil, por ejemplo, para la resolución de problemas referentes al encenagamiento del agua de un pantano, problemas cuya importancia es obvia en el momento actual de construcción de enormes estaciones hidroeléctricas.

Otro ejemplo, igualmente expresivo, lo ofrece la geometría no-euclídea¹, que apareció como culminación de una labor de dos mil años, iniciada en tiempos de Euclides para demostrar el axioma de las paralelas, un problema de interés puramente matemático. N. I. Lobachevski, fundador de la nueva geometría, tuvo cuidado en denominar a su geometría «imaginaria», puesto que no veía sentido para ella en el mundo real, aunque esperaba que algún día se encontraría. Los resultados de su geometría se les antojaban a la mayoría de los matemáticos no sólo «imaginarios», sino incluso inimaginables y absurdos. Sin embargo, sus ideas fueron el fundamento para un nuevo desarrollo de la geometría, con la aparición de teorías de espacios no-euclídeos; y posteriormente estas mismas ideas fueron la base de la teoría general de la relatividad, en la cual el aparato matemático consiste en una cierta forma de geometría no-euclídea de un espacio de cuatro dimensiones. Así, las construcciones abstractas de la matemática, que al principio parecieron incomprensibles, demostraron ser un poderoso instrumento para el desarrollo de una de las más importantes teorías de la física. Igual-

¹ Aquí nos limitamos a dar este ejemplo sin más explicaciones; para ello el lector puede consultar el capítulo 16.

mente, en la teoría actual de los fenómenos atómicos, en la llamada mecánica cuántica, es esencial el uso de numerosos conceptos matemáticos extremadamente abstractos, como por ejemplo el concepto de espacio de dimensión infinita.

No hay necesidad de dar otros ejemplos, puesto que ya hemos demostrado con suficiente énfasis que la matemática encuentra extensa aplicación en la vida diaria, en la tecnología y en la ciencia; en las ciencias exactas y en los problemas más complicados de la tecnología encuentran aplicación incluso aquellas teorías que nacen de la matemática misma. Esta es una de las características peculiares de la matemática, junto con su abstracción y el rigor y conclusión de sus resultados.

2. *Naturaleza esencial de la matemática*

Discutiendo estos rasgos especiales de la matemática estamos lejos de explicar su esencia; más bien hemos mostrado simplemente sus signos externos. Nuestra tarea ahora es explicar la naturaleza esencial de estos rasgos característicos. Para ello, será necesario responder, por lo menos, a las siguientes preguntas:

¿Qué reflejan estos conceptos matemáticos abstractos? En otras palabras, ¿cuál es el verdadero objeto de la matemática?

¿Por qué los resultados abstractos de la matemática parecen tan convincentes, y sus conceptos iniciales tan obvios? En otras palabras, ¿sobre qué cimientos reposan los métodos matemáticos?

¿Por qué, a pesar de toda su abstracción, encuentra la matemática tan amplias aplicaciones y no se queda simplemente en un juego fútil de abstracciones? En otras palabras, ¿cómo se explica el significado de la matemática?

Finalmente, ¿qué fuerzas llevan a nuevos desarrollos de la matemática, permitiendo unir la abstracción con la amplitud de sus aplicaciones? ¿Cuál es la base para su continuo crecimiento?

Resolviendo estas cuestiones formaremos un cuadro general del contenido de la matemática, de sus métodos, de su significado y su desarrollo; esto es, entenderemos su esencia.

Idealistas y metafísicos no sólo se ven sumidos en un mar de confusiones al intentar responder a estas cuestiones básicas, sino que llegan a distorsionar completamente la matemática, volviéndola literalmente del revés. Así, viendo la extrema abstracción y fuerza lógica de los resultados matemáticos, los idealistas imaginan que la matemática brota del pensamiento puro.

En realidad, la matemática no ofrece el más ligero soporte para

el idealismo o la metafísica. De esto nos convenceremos cuando intentemos resolver, en un esquema general, las cuestiones que acabamos de enunciar acerca de la esencia de la matemática. Para una clarificación preliminar de estas cuestiones, es suficiente examinar los fundamentos de la aritmética y la geometría elemental, a las cuales dirigimos ahora nuestra atención.

§ 2. *Aritmética*

1. *El concepto de número entero*

El concepto de número (por el momento hablaremos sólo de números enteros positivos), que tan familiar nos es hoy a nosotros, fue elaborado muy lentamente. Esto puede verse en el modo de contar de distintas razas que hasta tiempos muy recientes han permanecido en un nivel relativamente primitivo de vida social. En algunas de ellas los números mayores que dos o tres no tenían ya nombre; en otras llegaban algo más lejos pero terminaban al cabo de pocos números; para los restantes decían simplemente «muchos» o «incontables». Sólo gradualmente se fueron acumulando en los pueblos un conjunto de nombres claramente distintos para los números.

Al principio estos pueblos no tenían la noción de número, aunque podían, a su manera, juzgar sobre el tamaño de una u otra colección de objetos con los que se encontraban a diario. Debemos concluir que los números eran directamente percibidos por ellos como una propiedad inseparable de una colección de objetos, una propiedad que ellos, sin embargo, no podían claramente distinguir. Hoy día estamos tan acostumbrados a contar que difícilmente podemos imaginar este estado de cosas, pero es posible entenderlo ².

A un nivel inmediatamente superior, el número aparece ya como una propiedad de una colección de objetos, aunque no se distingue todavía de la colección en cuanto «número abstracto», en cuanto número no relacionado con objetos concretos. Esto es obvio si observamos los nombres que reciben algunos números entre ciertos

² De hecho, toda colección de objetos, tanto si es un rebaño de ovejas como un haz de leña, existe y es inmediatamente percibida en toda su concreción y complejidad. Distinguir en ella propiedades y relaciones es resultado de un análisis consciente. El pensamiento primitivo no realiza todavía este análisis, sino que considera los objetos como un todo. De modo semejante, un hombre que no ha estudiado música percibe una composición musical sin distinguir en ella detalles de la melodía, tonalidad, etc., mientras que un músico analiza fácilmente incluso una complicada sinfonía.

pueblos: «mano» para cinco y «hombre completo» para veinte. Aquí cinco se entiende no en sentido abstracto, sino simplemente en el sentido de «tantos como los dedos de una mano»; veinte es «tantos como los dedos de las manos y los pies de un hombre», y así sucesivamente. De un modo completamente análogo, ciertos pueblos no tenían los conceptos de «negro», «duro» o «circular». Para decir que un objeto es negro, lo comparaban con un cuervo, por ejemplo, y para decir que había cinco objetos, comparaban directamente estos objetos con una mano. De este modo también ocurrió que se utilizaron distintos nombres para un mismo número de objetos distintos; ciertos números para contar personas, otros para contar botes, y así sucesivamente, hasta llegar incluso a diez diferentes clases de números. Pero no se trata de números abstractos, sino simplemente de una especie de «apelación» referida sólo a una clase concreta de objetos. Otros pueblos no tenían en general nombres para designar los números; por ejemplo, no existía la palabra «tres», aunque podían decir «tres hombres» o «en tres lugares», etc.

Entre nosotros mismos ocurre algo parecido: a menudo decimos que este o aquel objeto es negro, pero pocas veces hablamos acerca de la «negrura» en sí misma, que es un concepto más abstracto³.

El número de objetos de una colección dada es una propiedad de la colección, pero el número en sí, el «número abstracto», es una propiedad abstraída de la colección concreta y considerada simplemente en sí misma, al igual que «negrura» o «dureza». Y lo mismo que la negrura es una propiedad común a todos los objetos del color del carbón, así el número «cinco» es la propiedad común a todas las colecciones que contienen objetos como dedos hay en una mano. En este caso la igualdad de los dos números se establece por simple comparación: tomamos un objeto de la colección, doblamos un dedo, y así hasta terminar la colección. En general, apareando los objetos de dos colecciones es posible establecer, sin hacer uso para nada de los números, si las colecciones tienen o no el mismo número de objetos. Por ejemplo, cuando los huéspedes pasan a ocupar sus luga-

³ En la formación de los conceptos referentes a propiedades de los objetos, tales como el color o la extensión de una colección, es posible distinguir tres etapas, que no se deben, naturalmente, intentar separar tajantemente unas de otras. En la primera etapa la propiedad se define comparando directamente los objetos: «como un cuervo», «tantos como en una mano». En la segunda aparece un adjetivo: «una piedra negra» o «cinco árboles» (el adjetivo numérico es muy análogo en este sentido). En la tercera etapa se abstrae la propiedad de los objetos y puede aparecer «como tal»; por ejemplo «negrura», o el número abstracto «5».

res en la mesa, la patrona sabe inmediatamente, sin necesidad de contar, si ha puesto o no un cubierto de menos, para lo cual basta mirar si algún comensal ha quedado sin lugar en la mesa.

De este modo es posible dar la siguiente definición: un número (tal como «dos», «cinco», etc.) es aquella propiedad de las colecciones de objetos que es común a todas las colecciones cuyos objetos pueden ponerse en correspondencia biunívoca unos con otros, y que es diferente en aquellas colecciones para las cuales tal correspondencia es imposible. Para descubrir esta propiedad y distinguirla claramente —esto es, para formar el concepto de número y darle un nombre: «seis», «diez», etc.— fue necesario comparar entre sí muchas colecciones de objetos. Durante generaciones y generaciones la gente repitió la misma operación millones de veces y de este modo descubrió los números y las relaciones entre ellos.

2. *Relaciones entre los números enteros*

Las operaciones con números aparecen como reflejo de las relaciones entre los objetos concretos. Esto se observa incluso en los nombres de los números. Por ejemplo, entre ciertos indios americanos el número «veintiséis» se pronuncia como «encima de dos dieces coloco un seis», que es claramente reflejo de un método concreto de contar objetos. La adición de números corresponde a situar juntas o unidas dos o más colecciones, y es igualmente fácil entender el significado concreto de la sustracción, multiplicación y división. La multiplicación en particular se debió en gran parte, como parece claro, al hábito de contar colecciones iguales: esto es, por doses, por treses, etc.

En el proceso de contar, los hombres no sólo descubrieron y asimilaron las relaciones entre los números, como, por ejemplo, que dos y tres son cinco, sino que también fueron estableciendo gradualmente ciertas leyes generales. Experimentalmente se descubrió que una suma no depende del orden de los sumandos y que el resultado de contar un conjunto dado de objetos no depende del orden en que se cuente, hecho que se refleja en la identidad esencial de los números «ordinal» y «cardinal»: primero, segundo, tercero, y uno, dos, tres. De este modo los números aparecen no como entidades separadas e independientes, sino relacionadas unas con otras.

Algunos números se expresan y se escriben en términos de otros. Así, en inglés «veinte» denota «dos (veces) diez»; en francés ochenta es «cuatro-veintes» (*quatre vingt*), noventa es «cuatro-veintes-diez»; y los números romanos VIII y IX, por ejemplo, denotan que $8 = 5 + 3$ y que $9 = 10 - 1$.

En general, los números no aparecieron como entidades separadas, sino como un sistema con sus relaciones mutuas y sus reglas.

El objeto de la aritmética es exactamente éste, el sistema de números con sus relaciones mutuas y sus reglas ⁴. Los números abstractos en sí no tienen propiedades tangibles y en general se puede decir muy poco sobre ellos. Si nos preguntamos, por ejemplo, por las propiedades del número seis, observamos que $6 = 5 + 1$, $6 = 3 \cdot 2$, que 6 es factor de 30, etc. Pero aquí el número 6 está siempre relacionado con otros números; de hecho, las propiedades de un número dado consisten precisamente en sus relaciones con otros números ⁵. Está claro, por consiguiente, que toda operación aritmética determina una conexión o relación entre números. Así, el objeto de la aritmética son las relaciones entre números. Pero estas relaciones son las imágenes abstractas de las relaciones cuantitativas reales entre colecciones de objetos; así podemos decir que la aritmética es la ciencia de las relaciones cuantitativas reales consideradas abstractamente, esto es, simplemente como relaciones. La aritmética, como vemos, no surge del pensamiento puro, como pretenden los idealistas, sino que es reflejo de propiedades definidas de las cosas reales; surge de una larga experiencia práctica de muchas generaciones.

3. *Símbolos numéricos*

A medida que la vida social se hizo más intensa y complicada, fueron apareciendo problemas más complejos. No sólo fue necesario anotar el número de objetos de un conjunto y comunicárselo a otros —necesidad que ya había conducido a la formulación del concepto de número y su denominación—, sino que llegó un momento en que fue esencial aprender a contar colecciones cada vez mayores de animales en un rebaño, de objetos para trueque, de días anteriores a una

⁴ La palabra «aritmética», que, significando «arte de calcular», deriva del adjetivo griego «aritmética», formado a partir del sustantivo «arithmos», que significa «número». El adjetivo modifica el nombre «techné» (arte, técnica), que aquí se sobreentiende.

⁵ Esto se deduce de consideraciones muy generales. Cualquier abstracción, eliminada su base concreta (igual que un número se abstrae de una colección completa de objetos), carece de sentido «en sí misma»; sólo existe en sus relaciones con otros conceptos. Estas relaciones ya están implícitas en cualquier afirmación sobre la abstracción, en su más imperfecta definición. Sin ellas la abstracción pierde todo contenido y significado, es decir, sencillamente no existe. El contenido del concepto de número abstracto reside en las reglas, en las relaciones mutuas del sistema de números.

fecha fijada, etc., y comunicar el resultado de la operación a otras personas. Esta situación pedía sin demora un perfeccionamiento en los nombres y símbolos de los números.

La introducción de los símbolos numéricos, que aparentemente se produjo al mismo tiempo que la escritura, jugó un gran papel en el desarrollo de la aritmética. Además fue la primera etapa hacia los signos matemáticos y las fórmulas en general. La segunda etapa, que consistió en la introducción de signos para las operaciones aritméticas y de una designación literal para la incógnita (x), tuvo lugar mucho más tarde.

El concepto de número, como el de cualquier otro concepto abstracto, no tiene una imagen inmediata; no puede ser exhibido, sino sólo concebido en la mente. Pero el pensamiento se formula en el lenguaje, y esto hace que sin nombres no pueda haber conceptos. El símbolo es también un nombre, excepto que no es oral sino escrito y se presenta a la mente en forma de una imagen visible. Por ejemplo, si digo «siete», ¿qué se imagina el lector? Probablemente no un conjunto de siete objetos de una u otra clase, sino más bien el símbolo «7», que forma una especie de marco tangible para el número abstracto «siete». Además, un número como 18273 es mucho más difícil de pronunciar que de escribir y no puede ser imaginado con ninguna precisión en forma de un conjunto de objetos. De este modo ocurrió (aunque sólo pasado cierto tiempo) que los símbolos dieron lugar a la concepción de números tan grandes que nunca habrían podido ser descubiertos por observación directa o por enumeración. Con la aparición del Estado surgió la necesidad de recoger impuestos, reclutar y equipar ejércitos, etc., todo lo cual requirió operaciones con números muy grandes.

Así, pues, la importancia de los símbolos numéricos reside en primer lugar en que suministran una materialización sencilla del concepto de número abstracto⁶. Y este es el papel de las notaciones matemáticas en general: conferir una estructura «tangible» a los conceptos matemáticos abstractos. Así, $+$ nota adición, x nota

⁶ Es importante señalar que el concepto de número, que fue elaborado con tanta dificultad a lo largo de tanto tiempo, lo domina hoy cualquier niño con relativa facilidad. ¿Por qué? La primera razón es, naturalmente, que el niño oye y ve a los adultos hacer constante uso de los números, e incluso le enseñan a hacer lo mismo. Pero una segunda razón, y a ésta es a la que deseamos prestar especial atención, es que el niño ya dispone de palabras y signos para los números. Aprende primero estos símbolos numéricos y sólo más tarde comprende su significado.

número desconocido, a un número cualquiera dado, etc. En segundo lugar, los símbolos numéricos proporcionan un medio particularmente sencillo de realizar operaciones con ellos. Todo el mundo sabe cuánto más fácil es «coger lápiz y papel» que «calcular de cabeza». Los signos matemáticos y las fórmulas tienen esta ventaja en general; permiten reemplazar una parte del razonamiento con cálculos, por algo que es casi mecánico. Además, los cálculos escritos poseen ya una autenticidad categórica; todo es visible, todo puede ser comprobado, y todo está definido por reglas exactas. Como ejemplos se podrían mencionar la adición por columnas o cualquier transformación algebraica tal como «pasar al otro miembro de la ecuación cambiando el signo». De lo dicho se desprende claramente que sin símbolos convenientes para los números la aritmética no habría podido hacer muchos progresos. Y lo que es más, las matemáticas contemporáneas serían imposibles sin la ayuda de signos especiales y fórmulas.

Es obvio que ese método sumamente conveniente de escribir los números que está hoy en uso no se pudo inventar de una vez. Desde los tiempos antiguos aparecieron en los distintos pueblos, en los comienzos de sus culturas, símbolos numéricos que eran muy diferentes de los actuales, no sólo en su apariencia general, sino también en los principios en que se fundaban.

Por ejemplo, el sistema decimal no se empleaba en absoluto y los antiguos babilonios tenían un sistema que era parcialmente decimal y parcialmente sexagesimal. La tabla 1 da algunos de los símbolos numéricos empleados por distintos pueblos. En particular vemos que los antiguos griegos, y más tarde también los rusos, hicieron uso de letras para designar los números. Nuestros actuales símbolos «arábigos», y en general nuestro método de formar los números, fueron traídos de la India a Europa por los árabes en el siglo x y arraigaron firmemente en el transcurso de pocos siglos.

La primera particularidad de nuestro sistema es que es un sistema decimal. Pero esto no es de gran importancia, puesto que habría sido posible usar, por ejemplo, un sistema duodecimal introduciendo símbolos especiales para diez y once. La particularidad más importante de nuestro sistema de designar números es que es «posicional»; esto es, un mismo dígito tiene distinto significado según sea su posición. Por ejemplo, en 372 el número 3 nota el número de centenas, y 7, el número de decenas. Este método de escritura no sólo es conciso y sencillo, sino que facilita grandemente el cálculo. Los numerales romanos eran a este respecto mucho menos manejables; el mismo

Tabla

SÍMBOLOS NUMÉRICOS EMPLEADOS

Eslavo			Chino			Griego
Cirílico	Glagolítico		Antiguo	Comercial	Científico	
0				○	○	
1	Ѡ	Ⲁ	一	丨	Ⅰ	α
2	Ѣ	ⲡ	二	𠄎	Ⅱ	β
3	Ѥ	Ⲣ	三	𠄎𠄎	Ⅲ	γ
4	Ѧ	ⲣ	四	𠄎𠄎𠄎	Ⅳ	δ
5	Ѩ	Ⲥ	五	𠄎𠄎𠄎𠄎	Ⅴ	ε
6	Ѭ	ⲥ	六	𠄎𠄎𠄎𠄎𠄎	Ⅵ	ς
7	Ѯ	Ⲧ	七	𠄎𠄎𠄎𠄎𠄎𠄎	Ⅶ	ζ
8	Ѱ	ⲧ	八	𠄎𠄎𠄎𠄎𠄎𠄎𠄎	Ⅷ	η
9	Ѳ	Ⲩ	九	𠄎𠄎𠄎𠄎𠄎𠄎𠄎𠄎	Ⅸ	θ
10	Ѵ	ⲩ	十	+	Ⅹ	ι
20	Ѷ	Ⲫ	二十	𠂇	Ⅺ	κ
30	Ѹ	ⲫ	三十	𠂇𠂇	Ⅻ	λ
100	Ѻ	Ⲭ	百	𠂇𠂇𠂇	ⅰ	ρ
1000	Ѽ	ⲭ	千	𠂇𠂇𠂇𠂇	ⅱ	α

1

POR LOS DISTINTOS PUEBLOS

Egipcio

Arabe	Georgiano	Jeroglífico	Hierático	Romano	Maya
0					
1 ١	Ⴕ	𐀀	𐦢	I	•
2 ٢	Ⴖ	𐀁	𐦣	II	••
3 ٣	Ⴗ	𐀂	𐦤	III	•••
4 ٤	Ⴘ	𐀃	𐦥	IV	••••
5 ٥	ႵႵ	𐀄	𐦦	V	—
6 ٦	ႵႶ	𐀅	𐦧	VI	—•
7 ٧	ႵႷ	𐀆	𐦨	VII	—••
8 ٨	ႵႸ	𐀇	𐦩	VIII	—•••
9 ٩	ႵႹ	𐀈	𐦪	IX	••••
10 ١٠	Ⴜ	𐀉	𐦫	X	==
20 ٢٠	ႼႵ	𐀊𐀉	𐦫𐦢	XX	
30 ٣٠	ႼႶ	𐀊𐀁	𐦫𐦣	XXX	
100 ١٠٠	ႼႷ	𐀊𐀂	𐦫𐦤	C	
1000 ١٠٠٠	ႼႸ	𐀊𐀃	𐦫𐦥	M	

número 372 se escribe CCCLXXII; además, es una tarea muy laboriosa multiplicar dos números elevados escritos en caracteres romanos.

La escritura posicional requiere que de un modo u otro se especifique que una cierta categoría de números ha sido omitida, puesto que de no hacerlo así confundiríamos, por ejemplo, treinta y uno con trescientos uno. En el lugar de la categoría omitida debemos colocar un cero, y de ese modo distinguimos 31 y 301. En forma rudimentaria, el cero ya aparece en las últimas escrituras cuneiformes babilónicas, pero su introducción sistemática fue obra de los indios⁷; ello les permitió elaborar un sistema de escritura completamente posicional como el que tenemos hoy día.

De esta forma, el cero llegó también a considerarse como un número y entró a formar parte del sistema de numeración. Pero en sí mismo el cero no es nada; en el lenguaje sánscrito de la antigua India se le llamaba exactamente eso: «vacío» (*çunga*); no obstante, en conexión con otros números el cero adquiere sentido y propiedades conocidas; por ejemplo, un número arbitrario más cero es el mismo número; un número arbitrario multiplicado por cero da cero.

4. *La teoría de números como una rama de la matemática pura*

Volvamos a la aritmética de los antiguos. Los textos más antiguos que se conservan de los tiempos de Babilonia y Egipto datan del segundo milenio a. C. Estos y otros textos posteriores contienen diversos problemas aritméticos con sus soluciones, entre ellos algunos que hoy pertenecen al álgebra, tales como la solución de ecuaciones cuadráticas e incluso cúbicas y progresiones; todo ello presentado, por supuesto, en forma de problemas concretos y ejemplos numéricos. Entre los babilonios también se encuentran tablas de cuadrados, cubos y recíprocos. Ello hace suponer que por entonces la matemática ya empezaba a suscitar cierto interés aparte del puramente relacionado con problemas prácticos.

En cualquier caso, la aritmética estaba bien desarrollada en la antigua Babilonia y Egipto, aun cuando no era todavía una teoría matemática de los números, sino más bien una colección de soluciones a ciertos problemas y de reglas de cálculo. Es exactamente de este modo como la aritmética se desarrolla actualmente en nues-

⁷ El primer manuscrito indio en el que figura el cero apareció a finales del siglo IX; en él el número 270 se escribe exactamente como hoy. Pero es probable que el cero fuese introducido en la India bastante antes, en el siglo VI.

tras escuelas primarias y es entendida por aquellos que no están especialmente interesados en la matemática. Lo cual es perfectamente legítimo, pero sin olvidar que la aritmética en esta forma no es todavía una teoría matemática al no haber teoremas generales sobre los números.

La transición a la aritmética teórica tuvo lugar gradualmente. Como ya hemos dicho, la existencia de los símbolos permite operar con números tan grandes que sería imposible visualizarlos como colecciones de objetos o llegar a ellos por el proceso de contar a partir del número uno. Las tribus primitivas tenían números especiales hasta 3, 10, 100, etc., pero después venía el indefinido «muchos». En contraste con esta situación, el uso de símbolos para los números permitió a los chinos, babilonios y egipcios llegar a las decenas de millares e incluso a los millones. Fue en esta etapa cuando se descubrió la posibilidad de prolongar indefinidamente la serie de los números, aunque no sabemos el momento en que esta posibilidad fue claramente percibida. Arquímedes (287-212 a. C.), por ejemplo, en su notable ensayo «El Arenario» se toma el trabajo de describir un método para designar un número mayor que el número de granos de arena suficiente para llenar «la esfera de las estrellas fijas». Esto indica que la posibilidad de nombrar y escribir tal número requería todavía en su tiempo una explicación detallada.

En el siglo III a. C., los griegos habían reconocido claramente dos ideas importantes: primera, que la sucesión de números era susceptible de ser prolongada indefinidamente; y segunda, que no sólo era posible operar con números cualesquiera dados, sino también referirse a los números en general y formular y probar teoremas sobre ellos. Esta idea representa la generalización de una cantidad inmensa de experiencias anteriores con números concretos, de las cuales entre-sacaron las reglas y métodos para razonamientos *generales* sobre los números. Se había operado una transición a un nivel más alto de abstracción: de números concretos (entes individuales, aunque abstractos) a números en general, es decir, a cualquier número posible.

A partir del sencillo proceso de contar objetos uno a uno, pasamos al proceso ilimitado de formación de números añadiendo una unidad al número anterior. La sucesión de números aparece como indefinidamente prolongable, y con ello entra en la matemática la noción de infinito. Naturalmente, en la práctica es imposible llegar, por el proceso de añadir una unidad tras otra, tan lejos como queramos en la sucesión de números: ¿quién podría alcanzar tanto como un millón de millones, que es casi cuarenta veces el número de segun-

dos que hay en mil años? Pero no es eso a lo que vamos; el proceso de añadir unidades, el proceso de formar colecciones arbitrariamente grandes de objetos es en principio ilimitado, por lo que existe la posibilidad de continuar la sucesión de los números más allá de todo límite. El hecho de que en la práctica real el proceso de contar sea limitado no hace aquí al caso. Es con esta sucesión indefinidamente prolongada con la que tienen que ver los teoremas generales sobre números.

Los teoremas generales sobre cualquier propiedad de un número arbitrario contienen ya en forma implícita infinidad de asertos sobre las propiedades de cualquier número y son por tanto cualitativamente más ricos que cualesquiera asertos particulares que pudieran verificarse para números específicos. Por esta razón los teoremas generales deben ser probados por razonamientos generales derivados de la regla fundamental de formación de la sucesión de los números. Aquí percibimos una profunda particularidad de la matemática: la matemática tiene como objeto no sólo relaciones cuantitativas dadas, sino todas las posibles relaciones cuantitativas, y entre ellas el infinito.

En los famosos «Elementos» de Euclides, escritos en el siglo III antes de C., ya encontramos teoremas generales sobre los números, en particular aquel que afirma que existen números primos infinitamente grandes ⁸.

Así es como la aritmética se transforma en la teoría de números: alejándose de los problemas concretos particulares para internarse en la región de los conceptos y razonamientos abstractos. Se ha convertido en una parte de la matemática «pura». En rigor este fue el momento del nacimiento de la matemática pura con los rasgos característicos discutidos en la primera sección. Debemos, naturalmente, anotar el hecho de que la matemática pura nació simultáneamente de la aritmética y de la geometría, y que ya habían sido encontrados en las reglas generales de la aritmética algunos de los rudimentos del álgebra, parte que se separó de la aritmética en una etapa posterior. Pero esto se discutirá más tarde.

Queda ahora resumir nuestras conclusiones hasta este punto, puesto que hemos trazado ya, aunque a grandes rasgos, el proceso por el cual la aritmética teórica surge a partir del concepto de número.

⁸ Recordemos que un número primo se define como un entero positivo mayor que la unidad que sólo es divisible por sí mismo y por la unidad.

5. *La naturaleza esencial de la aritmética*

Puesto que el nacimiento de la aritmética teórica es parte del nacimiento de la matemática, resulta razonable esperar que nuestras conclusiones acerca de la aritmética arrojen luz sobre nuestras primeras preguntas referidas a la matemática en general. Recordemos estas preguntas, particularmente en sus aplicaciones a la aritmética.

1. ¿Cómo surgieron los conceptos abstractos de la aritmética y qué reflejan en el mundo real?

Esta pregunta está respondida por las primeras observaciones acerca del nacimiento de la aritmética. Sus conceptos se corresponden con las relaciones cuantitativas de las colecciones de objetos. Estos conceptos surgen por la vía de la abstracción, como resultado del análisis y generalización de una inmensa cantidad de experiencia práctica. Aparecen gradualmente; primero aparecieron los números relacionados con objetos concretos, luego los números abstractos y finalmente el concepto de número en general, de cualquier número posible. Cada uno de estos conceptos surgió por combinación de la experiencia práctica y de conceptos abstractos anteriores. Esta es, de paso, una de las leyes de formación fundamentales de los conceptos matemáticos: los conceptos aparecen tras una serie de sucesivas abstracciones y generalizaciones, cada una de las cuales reposa en la combinación de experiencias con conceptos abstractos previos. La historia de los conceptos de la aritmética muestra cuán equivocado es el punto de vista idealista de que surgen del «pensamiento puro», de la «intuición innata», de la «contemplación de formas *a priori*», o algo similar.

2. ¿Por qué son las conclusiones de la aritmética tan convincentes e inalterables?

También la historia nos responde a esta pregunta. Ya vimos que las conclusiones de la aritmética fueron apareciendo lenta y gradualmente; reflejan la experiencia acumulada en el curso de incontables generaciones y de esta forma se han fijado indeleblemente en la mente del hombre. Dichas conclusiones también se han fijado en el lenguaje: en los nombres de los números, en sus símbolos, en la constante repetición de las mismas operaciones con los números, en su constante aplicación a la vida diaria. Es de este modo como han ido ganando en claridad y certidumbre. Los métodos de razonamiento lógico también tienen el mismo origen. Lo esencial no es sólo el hecho de que pueden repetirse cuando se desee, sino el de su validez y claridad, que poseen en común con las relaciones entre los seres

del mundo real, relaciones que se reflejan en los conceptos de la aritmética y en las reglas de la deducción lógica.

Esta es la razón por la cual los resultados de la aritmética son tan convincentes; sus conclusiones se siguen lógicamente de sus conceptos básicos; y unos y otros, los métodos de la lógica y los conceptos de la aritmética, fueron elaborados y fijados en nuestro conocimiento tras tres mil años de experiencia práctica, sobre la base de regularidades objetivas del mundo que nos rodea.

3. ¿Por qué tiene la aritmética tantas aplicaciones a pesar de la abstracción de sus conceptos?

La respuesta es sencilla. Los conceptos y conclusiones de la aritmética, que generalizan una enorme cantidad de experiencia, reflejan en forma abstracta aquellas relaciones del mundo real que se encuentran constantemente y en todas partes. Es posible contar los objetos de una habitación, las estrellas, la gente, los átomos, etc. La aritmética considera algunas de sus propiedades generales, haciendo abstracción de todo lo particular y concreto, y es precisamente porque se consideran únicamente estas propiedades generales por lo que sus conclusiones son aplicables a tantos casos. La posibilidad de un amplio rango de aplicaciones está garantizada por la gran abstracción de la aritmética, aunque es importante hacer notar que esta abstracción no es vacía, sino que se deriva de una gran experiencia práctica. Lo mismo es cierto para toda la matemática y para cualquier concepto abstracto o teoría. Las posibilidades de aplicación de una teoría dependen en gran medida del material original que ella generaliza.

Al mismo tiempo, todo concepto abstracto, y en particular el concepto de número, está limitado en su significado como resultado de su misma abstracción. En primer lugar, cuando es aplicado a cualquier objeto concreto refleja sólo un aspecto de éste y por tanto sólo proporciona una imagen incompleta de él. A menudo sucede, por ejemplo, que el simple dato numérico dice muy poco sobre la naturaleza del objeto. En segundo lugar, los conceptos abstractos no pueden aplicarse indiscriminadamente sino con ciertas limitaciones; es imposible aplicar la aritmética a problemas concretos sin asegurarnos primero de que su aplicación tiene sentido en ese caso particular. Si hablamos de adición, por ejemplo, y simplemente unimos los objetos en el pensamiento, entonces naturalmente no realizamos ningún proceso con los objetos mismos. Pero si aplicamos la adición a la reunión real de los objetos, si de hecho ponemos los objetos juntos —por ejemplo, apilándolos en un montón o colocándolos sobre una mesa—, entonces habremos efectuado una adición abs-

tracta, pero también un proceso real. Este proceso no consiste únicamente en una adición aritmética, que a veces puede ser incluso imposible de realizar. Por ejemplo, si arrojamus un objeto al montón, puede romperse; si se junta a animales salvajes de diversas especies, pueden despedazarse unos a otros; ciertos materiales puestos en contacto pueden reaccionar químicamente: un litro de agua y un litro de alcohol vertidos en un mismo recipiente producen no 2, sino 1,9 litros de mezcla, como resultado de la disolución parcial de los líquidos, etc. Ejemplos de este tipo son fáciles de encontrar.

Resumiendo, la realidad es concreta; y resulta particularmente importante recordar este hecho en relación con la matemática debido precisamente a su abstracción.

4. Finalmente, en la última cuestión nos preguntábamos por las fuerzas que condujeron al desarrollo de la matemática.

En el caso de la aritmética la respuesta también la proporciona la historia. Vimos cómo los pueblos aprendieron a contar y llegaron al concepto de número, y cómo las necesidades de la vida, planteando problemas más difíciles, requirieron la introducción de símbolos numéricos. En una palabra, las fuerzas que condujeron al desarrollo de la aritmética fueron las necesidades prácticas de la vida social. Estas necesidades prácticas y el pensamiento abstracto que surgió de ellas ejercieron unos sobre otros una constante interacción. Los conceptos abstractos constituyeron en sí una valiosa herramienta para la vida práctica y fueron constantemente mejorados debido a sus muchas aplicaciones. Al hacer abstracción de lo accidental se desvela lo esencial y garantiza el éxito en aquellos casos en que el papel importante corresponde precisamente a esas propiedades y relaciones elegidas y preservadas por la abstracción, y que son, en el caso de la aritmética, las relaciones cuantitativas.

Además, la reflexión abstracta a menudo va más lejos que las necesidades inmediatas de un problema práctico. Así, el concepto de números tan grandes como un millón o un billón surgió sobre la base de los cálculos prácticos, pero apareció antes que la necesidad práctica de hacer uso de ellos. En la historia de la ciencia abundan tales ejemplos; basta con recordar los números imaginarios que mencionamos anteriormente. Este es justamente un caso particular de un fenómeno conocido de todos: la interacción de la experiencia y el pensamiento abstracto, de la práctica y la teoría.

§3. Geometría

1. El concepto de figura geométrica

La historia del origen de la geometría es esencialmente similar al de la aritmética. Los conceptos geométricos más antiguos de los que se tiene noticia también pertenecen a los tiempos prehistóricos y son consecuencia igualmente de las actividades prácticas.

Los primeros hombres llegaron a las formas geométricas a través de la naturaleza. La luna llena y en cuarto creciente, la superficie lisa de un lago, la rectitud de un rayo de luz o de un árbol bien conformado existieron mucho antes que el hombre mismo, siendo desde el primer momento objeto de su observación. Pero en la naturaleza nuestros ojos raramente tropiezan con líneas auténticamente rectas, ni con triángulos o cuadrados perfectos, y es evidente que la principal razón por la cual el hombre logró gradualmente concebir estas figuras es la de que su observación de la naturaleza era activa, en el sentido de que para satisfacer sus necesidades prácticas manufacturaba objetos cada vez más regulares en su forma. Construía edificios, cortaba piedras, vallaba parcelas de tierra, tensaba cuerdas en sus arcos, modelaba cacharros de arcilla, llevándolos a la perfección y adquiriendo al mismo tiempo la noción de que una olla es *curva*, mientras que una cuerda tensa de arco es *recta*. En pocas palabras, el hombre primero dio forma a sus materiales y sólo más tarde reconoció la forma como algo que se imprime a la materia y que puede, por consiguiente, ser considerada en sí misma haciendo abstracción de aquélla. Reconociendo la forma de los cuerpos, el hombre logró mejorar su trabajo manual y de ese modo elaborar con mayor precisión aún la noción abstracta de forma. Fue así cómo las actividades prácticas sirvieron de base a los conceptos abstractos de la geometría. El hombre tuvo que manufacturar miles de objetos con bordes rectos, tensar miles de cuerdas, dibujar sobre el suelo un gran número de líneas rectas, antes de adquirir una noción clara de la línea recta en general como aquella cualidad común a todos estos casos particulares. En nuestros días el niño aprende bien pronto a dibujar una línea recta porque está rodeado de objetos con bordes rectos que son el resultado de una manufactura, y sólo por esta razón en nuestra infancia nos formamos ya una idea clara de la línea recta. Exactamente del mismo modo la noción de magnitudes geométricas, de longitud, área y volumen, surge de las actividades de la vida diaria. La gente medía longitudes, determinaba distancias, estimaba a ojo

el área de superficies y el volumen de los cuerpos, y todo ello por motivos prácticos. Fue así como se descubrieron las leyes generales más sencillas, las primeras relaciones geométricas: por ejemplo, que el área de un rectángulo es igual al producto de las longitudes de sus lados. Qué duda cabe que para un granjero es útil estar enterado de una relación que le permite estimar el área que tiene sembrada y consiguientemente la cosecha que puede esperar.

Vemos, pues, que la geometría tuvo su origen en las actividades prácticas y en los problemas de la vida cotidiana. Sobre esta cuestión escribió el sabio griego Eudemo de Rodas lo siguiente: «La geometría fue descubierta por los egipcios como resultado de las medidas de sus tierras, y estas medidas eran necesarias debido a las inundaciones del Nilo, que constantemente borran las fronteras⁹. No hay nada notable en el hecho de que esta ciencia, al igual que las otras, haya surgido de las necesidades prácticas del hombre. Todo conocimiento que surge de circunstancias imperfectas tiende por sí mismo a perfeccionarse. Surge de las impresiones de los sentidos, pero gradualmente se convierte en objeto de nuestra contemplación y finalmente entra en el reino del intelecto.»

Naturalmente, la medición de tierras no fue el único problema que llevó a los antiguos a la geometría. A partir de los textos fragmentarios que han llegado hasta nuestros días es posible formarse una idea de algunos de los problemas de los antiguos egipcios y babilonios y de sus métodos para resolverlos. Uno de los textos egipcios más antiguos es anterior al año 1700 a. C.; se trata de un manual de instrucción para «secretarios» (funcionarios reales), escrito por un tal Ahmes, y contiene una colección de problemas sobre cálculo de capacidades de contenedores y almacenes, de áreas de porciones de tierra, de dimensiones de terraplenes, etc.

Los egipcios y babilonios sabían determinar las áreas y los volúmenes más sencillos, conocían con considerable exactitud el cociente de la longitud de una circunferencia a su diámetro, y quizá incluso supieran calcular el área de la superficie de una esfera; en una palabra, poseían ya una cantidad respetable de conocimientos geométricos. Pero, por lo que hasta ahora sabemos, no estaban todavía en posesión de la geometría como ciencia teórica provista de teoremas y demostraciones. Al igual que la aritmética en aquel tiempo, la geometría era fundamentalmente una colección de reglas deducidas

⁹ Aquí se entiende límites entre terrenos colindantes. Señalemos, entre paréntesis, que *geometría* significa medida de tierras (en griego antiguo «ge» es tierra, y «metron», medida).

de la experiencia. Además, la geometría no se distinguía en general de la aritmética. Los problemas geométricos eran al mismo tiempo problemas de cálculo aritmético.

En el siglo VII a. C., la geometría pasó de Egipto a Grecia, donde continuó desarrollándose bajo la tutela de los grandes filósofos materialistas: Tales, Demócrito y otros. Los sucesores de Pitágoras, fundadores de una escuela religioso-filosófica idealista, hicieron también importantes contribuciones a la geometría.

El desarrollo de la geometría se vio encauzado hacia la recopilación de nuevos hechos y la clarificación de las relaciones de unos con otros. Estas relaciones se fueron transformando gradualmente en deducciones lógicas de unas proposiciones de la geometría a partir de otras, lo cual llevó a dos resultados: primero, al concepto de teorema geométrico y de su demostración; y segundo, a la clarificación de aquellas proposiciones fundamentales a partir de las cuales se pueden deducir las restantes, es decir, los axiomas.

De este modo la geometría se convirtió gradualmente en una teoría matemática.

Es bien sabido que las exposiciones sistemáticas de la geometría aparecieron en Grecia ya en el siglo V a. C., pero no se han conservado por la razón obvia de que todas fueron suplantadas por los «Elementos» de Euclides (siglo III a. C.). En este trabajo, la geometría se presentó como un sistema tan bien construido que sus fundamentos no sufrieron ninguna alteración esencial hasta llegar a N. I. Lobachevski, más de dos mil años después. El conocido texto escolar de Kiselev en sus primeras ediciones, al igual que los libros de texto de todo el mundo, no era sino una reelaboración de la obra de Euclides. Muy pocos libros en el mundo han tenido una vida tan larga como los «Elementos» de Euclides, perfecta creación del genio griego. Naturalmente, la matemática continuó avanzando, y nuestro conocimiento de los fundamentos de la geometría fue mejorando considerablemente; sin embargo, los «Elementos» de Euclides se convirtieron, y en gran parte siguen siendo, en modelo de un libro de matemática pura. Reuniendo los hallazgos de sus predecesores, Euclides presentó la matemática de su tiempo como una ciencia teórica independiente, esto es, la presentó en esencia tal como hoy la entendemos.

2. *La naturaleza esencial de la geometría*

La historia de la geometría lleva a las mismas conclusiones que la de la aritmética. Vemos que la geometría surge de la vida práctica

y que su transformación en teoría matemática requirió un inmenso período de tiempo.

La geometría opera con «cuerpos geométricos» y figuras; estudia sus relaciones mutuas desde el punto de vista de la magnitud y la posición. Pero un cuerpo geométrico no es sino un cuerpo real considerado únicamente desde el punto de vista de su forma espacial¹⁰ y haciendo abstracción de todas sus otras propiedades, tales como densidad, color o peso. Una figura geométrica es un concepto todavía más general, puesto que en este caso es posible abstraer también la extensión espacial; así una superficie tiene sólo dos dimensiones, una línea sólo una dimensión, y un punto ninguna. El punto es el concepto abstracto de final de una línea, de una posición definida con un máximo de precisión, de suerte que no se componga ya de parte alguna. Así es como fueron definidos por Euclides todos estos conceptos.

La geometría tiene, pues, como objeto las formas espaciales y las relaciones de los cuerpos reales, eliminando de ellos las restantes propiedades, y considerándolos desde un punto de vista puramente abstracto. Es justamente este alto nivel de abstracción el que distingue la geometría de las otras ciencias que también se ocupan de las formas espaciales y las relaciones de los cuerpos. En astronomía, por ejemplo, se estudian las posiciones mutuas de ciertos cuerpos, pero estos cuerpos son los cuerpos celestes reales; en geodesia es la forma de la Tierra la que se estudia; en cristalografía, la forma de los cristales, etcétera. En todas estas otras ciencias, la forma y posición de los cuerpos concretos se estudian en función de otras propiedades suyas.

Esta abstracción conduce necesariamente al método geométrico puramente teórico, pues como no es posible realizar experimentos con líneas rectas sin anchura con «formas puras», la única posibilidad es hacer uso de razonamientos lógicos, derivando unas conclusiones de otras. Un teorema geométrico debe ser probado mediante razonamientos, pues de otro modo, al no operar con «formas puras», no pertenece a la geometría.

La evidencia de los conceptos básicos de la geometría, sus métodos de razonamiento y la certeza de sus conclusiones tienen el mismo origen que en el caso de la aritmética. Las propiedades de los conceptos geométricos, al igual que los conceptos mismos, han sido abstraídos del mundo que nos rodea. Fue preciso que los hombres dibujaran muchas líneas rectas antes de que llegaran a considerar

¹⁰ Por forma entendemos también dimensiones.

como axioma que a través de dos puntos cualesquiera distintos es siempre posible dibujar una línea recta; y fue necesario mover muchos cuerpos y acoplarlos unos a otros incontables veces antes de que se pudiera pasar de esta experiencia a la noción de superposición de figuras geométricas y hacer uso de esta noción para la demostración de teoremas, como se hace en los conocidos teoremas sobre congruencia de triángulos.

Finalmente, debemos llamar la atención acerca de la generalidad de la geometría. El volumen de una esfera es igual a $\frac{4}{3}\pi r^3$ con independencia de si se trata de un recipiente esférico, de una esfera de acero, de una estrella, o de una gota de agua. La geometría puede abstraer lo que es común a todos los cuerpos, porque todo cuerpo real tiene una forma, unas dimensiones y una posición con respecto a los demás cuerpos más o menos definidas. No es extraño por tanto que la geometría tenga un campo de aplicaciones casi tan amplio como la aritmética. Los obreros que miden las dimensiones de un edificio o interpretan un plano, un artillero que determina la distancia al blanco, un granjero que mide la superficie de su granja, un ingeniero que estima el volumen de una infraestructura, todos ellos utilizan la geometría. El piloto, el astrónomo, el viajero, el ingeniero y el físico necesitan trabajar con las conclusiones precisas de la geometría.

Un claro ejemplo de la resolución geométrico-abstracta de un importante problema de física lo proporcionan las investigaciones del conocido cristalógrafo y geómetra E. S. Fedorov. El problema que se propuso a sí mismo de encontrar todas las formas posibles de simetría de los cristales es uno de los más fundamentales de la cristalografía teórica. Para resolverlo, Fedorov hizo abstracción de todas las propiedades físicas de un cristal, considerándolo sólo como un sistema regular de cuerpos geométricos «en lugar de un sistema de átomos concretos». Así el problema se convirtió en el de encontrar todas las formas de simetría que pudieran existir en un sistema de cuerpos geométricos. Este problema, puramente geométrico, fue completamente resuelto por Fedorov, hallando que existían 230 formas posibles de simetría. La solución resultó ser una contribución importante a la geometría y dio origen a gran número de investigaciones geométricas.

En este ejemplo, como en toda la historia de la geometría, observamos la fuerza principal que actúa en el desarrollo de la geometría. Es la influencia mutua de la vida práctica y el pensamiento abstracto. El problema de descubrir posibles simetrías, originado por la observación física de los cristales, se transformó en un problema abstracto

y dio lugar a una nueva teoría matemática, la teoría de los sistemas regulares, o de los llamados grupos de Fedorov¹¹. Más tarde esta teoría no sólo encontró brillante confirmación en la observación de los cristales, sino que también sirvió como guía general en el desarrollo de la cristalografía, dando lugar a nuevas investigaciones, tanto en física experimental como en la matemática pura.

§ 4. *Aritmética y geometría*

1. *El origen de las fracciones en la interrelación de la aritmética y la geometría*

Hasta ahora hemos considerado la aritmética y la geometría por separado. Su relación mutua, y por consiguiente la interrelación general de todas las teorías matemáticas, ha escapado hasta ahora a nuestra atención. Sin embargo esta relación tiene una importancia excepcional. La interrelación de las teorías matemáticas conduce a avances en la matemática en sí y también descubre un rico caudal de relaciones mutuas en el mundo real que reflejan estas teorías.

La aritmética y la geometría no sólo se aplican una a la otra, sino también son fuente de otros métodos, ideas y teorías generales. En último término, la aritmética y la geometría son las dos raíces sobre las cuales ha crecido toda la matemática. Su influencia mutua se hace sentir desde el mismo momento de su nacimiento. Incluso la simple medición de una línea representa una fusión de la geometría y la aritmética. Para medir la longitud de un objeto se le *aplica* a éste una cierta unidad de longitud y se *calcula* cuántas veces es posible repetir esa operación; el primer paso (aplicación) es de carácter geométrico, el segundo (cálculo), de carácter aritmético. Quien cuenta sus pasos al andar está ya uniendo estas dos operaciones.

En general, la medición de cualquier magnitud combina el cálculo con alguna operación específica que es característica de esta magnitud. Baste mencionar la medida de un líquido en un recipiente graduado o la de un intervalo de tiempo contando el número de oscilaciones de un péndulo.

Pero en el proceso de medida generalmente ocurre que la unidad elegida no está contenida un número entero de veces en la magnitud a medir, por lo que el simple cálculo del número de unidades no es suficiente. Surge entonces la necesidad de fraccionar la unidad de medida para poder expresar la magnitud con mayor exactitud en

¹¹ Compárese con el capítulo 19.

partes de la unidad; esto es, no mediante números enteros sino por medio de fracciones. Fue así como surgieron realmente las fracciones, hecho que se ha demostrado por el análisis de datos históricos y de otro tipo. Surgieron de la división y comparación de las magnitudes continuas; en otras palabras, de las mediciones. Las primeras magnitudes que se midieron fueron de carácter geométrico: longitudes, superficies de labranza y volúmenes de líquidos o de materiales desmenuzables, por lo que ya en la primera aparición de las fracciones se observa la acción mutua de la aritmética y la geometría. Esta interacción conduce a la aparición de un nuevo concepto importante, el de las fracciones, como extensión del concepto de número de los enteros a los fraccionarios (o, como dicen los matemáticos, a los racionales, expresados como cocientes de números enteros). Las fracciones no surgen, ni podrían surgir, de la división de números enteros, puesto que con números enteros sólo se cuentan objetos enteros. Tres hombres, tres flechas, etc., son cosas que tienen sentido; no así conceptos como dos tercios de un hombre e incluso dos tercios de una flecha; tres tercios de una flecha tomados por separado no matarían ningún ciervo, pues para ello se necesita una flecha *entera*.

2. *Magnitudes inconmensurables*

En el desarrollo del concepto de número, que surge de la acción mutua de la aritmética y la geometría, la aparición de las fracciones fue sólo la primera etapa. La siguiente fue el descubrimiento de los intervalos inconmensurables. Recordemos que dos intervalos se llaman inconmensurables si no existe ningún intervalo que pueda aplicarse a cada uno de ellos un número entero de veces, o, dicho con otras palabras, si su cociente no puede expresarse por una fracción ordinaria (un cociente de números enteros).

Al principio la gente sencillamente no pensó en la cuestión de si todo intervalo podía ser expresado por una fracción. Si al dividir o medir un intervalo se llegaba a partes muy pequeñas, éstas se desechaban sin más; en la práctica, no tenía sentido hablar de una precisión infinita en el proceso de medida. Demócrito llegó incluso a sugerir la idea de que las figuras geométricas constan de átomos de una clase particular. Esta noción, que para nosotros parece bastante extraña, demostró ser muy útil en la determinación de áreas y volúmenes. Las áreas se calculaban como la suma de filas formadas por átomos, y los volúmenes, como suma de capas de átomos. Fue así, por ejemplo, como Demócrito encontró el volumen de un cono. El lector familiarizado con el cálculo integral observará que este

método constituye ya el prototipo de la determinación de áreas o volúmenes por los métodos del cálculo integral. Por otra parte, al volver con el pensamiento a los tiempos de Demócrito debemos prescindir de las nociones hoy en día en uso, nociones que se han ido fijando firmemente en nuestras mentes como consecuencia del desarrollo de la matemática. En la época de Demócrito, las figuras geométricas no estaban todavía separadas de las reales tanto como ahora. Y puesto que Demócrito consideraba los cuerpos reales como compuestos de átomos, nada más natural que también contemplara las figuras geométricas bajo este mismo ángulo.

Pero la afirmación de que los intervalos constan de átomos está en contradicción con el teorema de Pitágoras, toda vez que de este teorema se sigue la existencia de intervalos inconmensurables. Por ejemplo, la diagonal de un cuadrado es inconmensurable con su lado; en otras palabras, el cociente de los dos no puede expresarse como cociente de números enteros.

Demostremos que el lado y la diagonal de un cuadrado son, de hecho, inconmensurables. Si a es el lado y b la diagonal de un cuadrado, entonces según el teorema de Pitágoras $b^2 = a^2 + a^2 = 2a^2$, y por tanto

$$\left(\frac{b}{a}\right)^2 = 2.$$

Ahora bien, no existe ninguna fracción tal que su cuadrado sea igual a 2. En efecto, supongamos que sí existe, y sean p y q números enteros para los cuales

$$\left(\frac{p}{q}\right)^2 = 2,$$

donde podemos suponer que p y q no tienen factores comunes, porque en otro caso podríamos simplificar la fracción. Pero si $(p/q)^2 = 2$, entonces $p^2 = 2q^2$, y por tanto p^2 es divisible por 2. En este caso p^2 es también divisible por 4, puesto que es el cuadrado de un número par. Así $p^2 = 4q_1$; esto es, $2q^2 = 4q_1$, y $q^2 = 2q_1$. De esto se sigue que q debe también ser divisible por 2. Pero esto contradice la suposición de que p y q no tienen factores comunes. Esta contradicción demuestra que el cociente b/a no puede ser expresado mediante un número racional. La diagonal y el lado de un cuadrado son inconmensurables.

Este descubrimiento produjo gran impresión a los científicos griegos. En nuestros días, acostumbrados como estamos a los números irracionales y operando cómodamente con raíces cuadradas, la exis-

tencia de intervalos inconmensurables no nos preocupa. Pero en el siglo v a. C., el descubrimiento de tales intervalos tuvo un aspecto completamente diferente para los griegos. Pues al no conocer el concepto de número irracional ni disponer de un símbolo como $\sqrt{2}$, el resultado anterior indicaba que el cociente de la diagonal y el lado del cuadrado no se podía representar por ningún número.

En la existencia de intervalos inconmensurables los griegos descubrieron una profunda paradoja inherente al concepto de continuidad, una de las expresiones de la contradicción dialéctica entre continuidad y movimiento. Muchos fueron los filósofos griegos que consideraron esta contradicción; particularmente conocido entre ellos, por sus paradojas, es Zenón de Elea.

Los griegos fundaron una teoría de cocientes de intervalos, o de magnitudes en general, que tiene en cuenta la existencia de los intervalos inconmensurables¹²; está expresada en los «Elementos» de Euclides, y en forma simplificada se explica hoy en los cursos de geometría del bachillerato. Pero a la idea de que el cociente entre dos intervalos (si el segundo intervalo se toma como unidad de longitud, este cociente es simplemente la longitud del primer intervalo) puede también ser considerado como un número (generalizando así el concepto mismo de número) no lograron acceder los griegos. El concepto de número irracional no se originó entre ellos¹³, sino que fue una aportación posterior de los matemáticos orientales; en realidad, una definición matemáticamente rigurosa de número real, no dependiente de manera directa de la geometría, sólo ha sido dada recientemente: en los años setenta del siglo pasado¹⁴. El hecho de que tuviese que

¹² Esta teoría se atribuye al científico griego Eudoxio, que vivió en el siglo iv a. C.

¹³ Como resultado del hecho de que la teoría de la medida de magnitudes no pasó a formar parte de la aritmética sino de la geometría, la matemática entre los griegos fue, sobre todo, geometría. Cuestiones tales como, por ejemplo, la solución de las ecuaciones de segundo grado, que hoy se estudian por métodos algebraicos, las plantearon y resolvieron geoméricamente. Los «Elementos» de Euclides contienen un número considerable de tales cuestiones, que representaron evidentemente para los matemáticos contemporáneos un resumen de los fundamentos, no sólo de la geometría en el sentido actual, sino de la matemática en general. Esta preponderancia de la geometría continuó hasta que Descartes, por el contrario, subordinó la geometría al álgebra. Se conservan trazas del largo dominio de la geometría en nombres tales como «cuadrado» y «cubo» para la segunda y tercera potencias: « a al cubo» es un cubo de lado a .

¹⁴ No estamos hablando aquí de una definición descriptiva, sino de una definición que sirva de base inmediata para la demostración de teoremas sobre propiedades de los números reales. Es natural que tales definiciones aparecieran

transcurrir un período tan largo de tiempo después de ser establecida la teoría de los cocientes demuestra lo difícil que es descubrir conceptos abstractos y darles una formulación exacta.

3. *El número real*

Al describir el concepto de número real, Newton, en su «Aritmética General», escribió: «por número entendemos no tanto una colección de unidades como un cociente abstracto de una cierta magnitud a otra tomada como unidad». Este número (cociente), puede ser entero, racional, o, si la magnitud dada es inconmensurable con la unidad, irracional.

Así, pues, el número real en su sentido original no es otra cosa que el cociente de una magnitud a otra tomada como unidad; en casos particulares es un cociente de intervalos, pero puede también ser un cociente de áreas, pesos, etc.

Por tanto, un número real es, en general, un cociente de magnitudes, en las cuales se ha hecho abstracción de su naturaleza concreta.

Igual que los números enteros *abstractos* sólo tienen interés matemático en sus relaciones mutuas, así los números reales *abstractos* sólo adquieren un contenido y se convierten en objeto de interés matemático cuando se relacionan entre sí dentro del sistema de los números reales.

En esta teoría, como en la de la aritmética, es necesario primero definir operaciones con los números: adición, substracción, multiplicación, división, así como las relaciones expresadas por palabras tales como «mayor que» o «menor que». Estas operaciones y relaciones reflejan conexiones reales entre las distintas magnitudes; por ejemplo, la adición refleja la colocación consecutiva de intervalos. Las primeras operaciones con números abstractos datan de la Edad Media y se deben a los matemáticos orientales. Más tarde vino el descubrimiento gradual de la propiedad más importante del sistema de los números reales: su continuidad. El sistema de los números reales es la imagen abstracta de todos los valores posibles de una magnitud que varía continuamente.

De este modo, y al igual que en el caso de los números enteros, la aritmética de los números reales consiste en las relaciones cuanti-

en un período posterior, cuando el desarrollo de la matemática, y en particular del análisis infinitesimal, necesitó de una definición conveniente de número real representado por «la variable x ». Esta definición fue dada en distintas formas durante los años setenta del siglo pasado por los matemáticos alemanes Weierstrass, Dedekind y Cantor.

tativas entre magnitudes continuas, estudiándolas en su forma general y haciendo completa abstracción de todas sus propiedades concretas. Es precisamente debido a que los números reales representan lo que es común a todas las magnitudes continuas por lo que tienen una aplicación tan amplia: los valores de diversas magnitudes —una longitud, un peso, la intensidad de una corriente eléctrica, la energía, etcétera— se expresan por números, y las interdependencias o relaciones entre estas magnitudes se traducen en relaciones entre sus valores numéricos.

Para demostrar cómo el concepto general de número real puede servir como base de una teoría matemática, es preciso dar su definición matemática de un modo formal. Esto puede hacerse por diferentes métodos, pero quizá el más natural es proceder a partir del proceso mismo de medida de magnitudes, que es realmente lo que condujo a esta generalización del concepto de número. Aunque aquí hablaremos de longitudes de intervalos, el lector se dará cuenta de que podríamos argüir exactamente de la misma forma con cualesquiera otras magnitudes que permitan una subdivisión indefinida.

Supongamos que deseamos medir el intervalo AB por medio del intervalo CD tomado como unidad (figura 1).

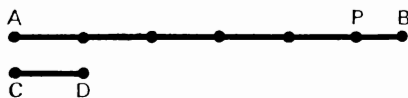


FIG. 1

Aplicamos el intervalo CD al AB (empezando por ejemplo por el punto A) tantas veces como CD esté contenido en AB . Supongamos que esto ocurre n_0 veces. Si todavía queda un resto PB del intervalo AB , entonces dividimos el intervalo CD en diez partes y medimos el resto con estas décimas. Supongamos que son n_1 décimas las que caben en el resto. Si después de ello queda todavía un resto, dividimos de nuevo la medida en diez partes, esto es, dividimos CD en cien partes, y repetimos la misma operación; y así sucesivamente. Una de dos: o el proceso de medida termina, o continúa; pero en cualquiera de los dos casos llegamos al resultado de que en el intervalo AB el intervalo CD está contenido n_0 veces, las décimas n_1 veces, las centésimas n_2 veces, etc. En una palabra, obtenemos el cociente de AB a CD con una aproximación creciente: hasta las décimas, hasta las centé-

simas, etc. Así el cociente podrá representarse por una fracción decimal con n_0 unidades, n_1 décimas, etc.:

$$\frac{AB}{CD} = n_0 \cdot n_1 n_2 n_3 \dots$$

Esta fracción decimal puede ser infinita, caso que corresponde a la posibilidad de un crecimiento indefinido en la precisión de la medida.

Así, pues, el cociente de dos intervalos, o de dos magnitudes en general, es siempre representable mediante una fracción decimal, finita o infinita. Pero en la fracción decimal no queda traza alguna de la magnitud concreta; aquélla representa exactamente el cociente abstracto, el número real. Así, un número real se puede definir formalmente, si lo deseamos, como una fracción decimal finita o infinita¹⁵.

Esta definición estará completa si decimos lo que entendemos por las operaciones de adición, etc., para las fracciones decimales. Esto se hace de tal modo que las operaciones definidas sobre las fracciones decimales correspondan a operaciones sobre las propias magnitudes. Así, cuando se colocan varios intervalos uno detrás de otro sus longitudes se suman; esto es, la longitud del intervalo $AB + BC$ es igual a la suma de las longitudes de AB y de BC . Al definir las operaciones sobre los números reales, existe la dificultad de que estos números vienen representados en general por fracciones decimales *infinitas*, mientras que las conocidas reglas para estas operaciones se refieren a fracciones decimales finitas. Sin embargo, puede darse una definición rigurosa de las operaciones para decimales infinitos del modo siguiente. Supongamos, por ejemplo, que queremos sumar dos números, a y b . Tomamos las correspondientes fracciones decimales hasta un lugar decimal dado, por ejemplo las millonésimas, y las sumamos, obteniendo así la suma $a + b$ con una aproximación que en este caso es de dos millonésimas, puesto que los errores en a y b se suman. De este modo podemos definir la suma de dos números *con cualquier grado de precisión que se desee*, y en este sentido su suma está completamente definida, aunque en cada etapa del cálculo se conozca sólo con una exactitud limitada. Ahora bien, este hecho se corresponde con la naturaleza esencial del caso,

¹⁵ Las fracciones con el nueve como dígito periódico no se consideran aquí; según una conocida regla, son idénticas a las fracciones correspondientes sin nueves, como en el ejemplo siguiente: $0,139999999\dots = 0,140000000\dots$

puesto que cada una de las magnitudes a y b se mide también sólo con una precisión limitada, y el valor exacto de cada una de las correspondientes fracciones infinitas se obtiene como resultado de un aumento indefinido en la aproximación; las relaciones «mayor que» y «menor que» pueden ser entonces definidas por medio de la adición: a es mayor que b si existe un número c tal que $a = b + c$, donde estamos hablando, naturalmente, de números positivos.

La continuidad de la sucesión de los números reales encuentra su expresión en el hecho de que, si los números $a_1, a_2, \dots$ van en aumento, y los $b_1, b_2, \dots$, en disminución, pero siempre permaneciendo superiores a los a_i , entonces entre la primera serie de números y la segunda hay siempre un número c . Esto puede visualizarse sobre una recta estableciendo una correspondencia entre los puntos de ésta y los números (Fig. 2), según la regla usual.

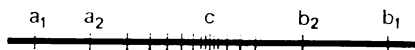


FIG. 2

Se ve claramente que la presencia del número c y del punto que le corresponde significa la ausencia de un corte en las series de números y esto es lo que se entiende por continuidad.

4. *El conflicto de los opuestos: lo concreto y lo abstracto*

En el ejemplo de la interacción de la aritmética con la geometría vimos ya que el desarrollo de la matemática es un proceso conflictivo entre muchos elementos contrastantes: lo concreto y lo abstracto, lo particular y lo general, lo formal y lo material, lo finito y lo infinito, lo discreto y lo continuo, etc. Tratemos, por ejemplo, de analizar el contraste entre lo concreto y lo abstracto en la formación del concepto de número real. Como hemos visto, los números reales reflejan un proceso de medida infinitamente mejorable o, en términos ligeramente diferentes, la determinación de una magnitud con absoluta exactitud. Esto se corresponde con el hecho de que en geometría consideramos formas y dimensiones de los cuerpos idealmente precisas, abstrayéndolas de la movilidad de los objetos concretos y de una cierta indefinibilidad de sus formas y dimensiones reales; por ejemplo, el intervalo que medimos en la figura 1 era completamente ideal.

Pero las formas geométricas idealmente precisas y los valores absolutamente precisos de las magnitudes no son sino abstracciones. Ningún objeto concreto tiene una forma absolutamente precisa ni ninguna magnitud concreta puede medirse con exactitud absoluta, puesto que ni siquiera *tiene* un valor perfectamente definido. La longitud de un segmento, por ejemplo, no tiene sentido si se trata de precisarla por encima de los límites de las dimensiones atómicas. Siempre que se superan ciertos límites bien conocidos de la precisión cuantitativa, aparece un cambio cualitativo en la magnitud, y ésta, en general, pierde su sentido original. Por ejemplo, la presión de un gas no puede precisarse más allá de los límites del impacto de una única molécula; la carga eléctrica deja de ser continua cuando se trata de medirla con una precisión mayor que la carga de un electrón, etc. En vista de que en la naturaleza no existen objetos de forma idealmente precisa, la afirmación de que el cociente de la diagonal de un cuadrado al lado es igual a $\sqrt{2}$ no sólo no puede deducirse con exactitud absoluta a partir de medidas concretas, sino que carece por completo de un significado preciso en el caso de un cuadrado concreto.

La conclusión de que la diagonal y el lado de un cuadrado son inconmensurables se sigue, como hemos visto, del teorema de Pitágoras. Se trata de una conclusión *teórica* basada en un desarrollo de los datos de la experiencia; es el resultado de la aplicación de la lógica a las premisas originales de la geometría, premisas que fueron tomadas de la experiencia.

De este modo, el concepto de los intervalos inconmensurables, y con más razón aún el de los números reales, no es simple reflejo inmediato de los hechos de la experiencia, sino que los trasciende. Hecho muy comprensible, pues el número real no refleja una magnitud concreta dada sino más bien la magnitud en general, con abstracción de toda su concreción; en otras palabras, refleja lo que es *común* a las magnitudes concretas particulares. Lo que es común a todas ellas consiste en particular en esto, en que el valor de la magnitud puede ser determinado con una precisión ilimitada; y si hacemos abstracción de las magnitudes concretas, entonces el límite de este posible aumento en la precisión, que depende de la naturaleza concreta de la magnitud, se hace indefinido y desaparece.

Así, pues, una teoría *matemática* de las magnitudes, desde el momento en que las considera haciendo abstracción de su naturaleza individual, tiene inevitablemente que considerar la posibilidad de una precisión ilimitada para el valor de la magnitud y *debe* por tanto

conducir al concepto de número real. Al mismo tiempo, puesto que refleja sólo lo que es común a varias magnitudes, la matemática no tiene en cuenta las peculiaridades de cada magnitud individual.

Dado que la matemática sólo somete a estudio propiedades generales, al operar con abstracciones perfectamente definidas lo hace con independencia de los límites reales de su aplicabilidad; y esto no podía ser de otra forma, ya que estos límites son distintos en cada caso particular, o lo que es lo mismo, dependen de las propiedades concretas de los fenómenos en consideración y de los cambios cualitativos que tienen lugar en ellos. Así, al aplicar la matemática a un caso concreto, es necesario verificar la aplicabilidad real de la teoría en cuestión. Considerar la materia como continua y describir sus propiedades mediante magnitudes continuas, sólo es posible si podemos hacer abstracción de su estructura atómica, lo cual sólo sucede en determinadas condiciones.

Sin embargo, los números reales representan un instrumento seguro y poderoso para la investigación matemática de magnitudes y procesos realmente continuos. Su teoría se basa en la práctica, en un inmenso campo de aplicaciones a la física, tecnología y química. Por consiguiente, la experiencia demuestra que el concepto de número real refleja fielmente las propiedades generales de las magnitudes. Pero esta fidelidad no deja de tener límites; no es lícito considerar la teoría de los números reales como algo absoluto y que permite un desarrollo abstracto ilimitado y divorciado por completo de la realidad. El concepto mismo de número real continúa desarrollándose y está de hecho muy lejos todavía de ser un concepto completo.

5. *El conflicto de los opuestos: lo discreto y lo continuo*

El papel que desempeña otro de los mencionados contrastes —el que existe entre lo discreto y lo continuo— puede ilustrarse también mediante el desarrollo del concepto de número. Ya hemos visto que las fracciones surgieron de la división de magnitudes continuas.

Sobre este tema de la división hay una anécdota extraordinariamente instructiva. La abuela ha comprado tres patatas y tiene que dividir las a partes iguales entre sus dos nietos. ¿Cómo debe hacerlo? La solución es: haciéndolas puré.

La anécdota revela la verdadera esencia de la materia. Cada objeto por separado es indivisible en el sentido de que, cuando se divide, casi siempre deja de ser lo que era, como se ve claramente en el ejemplo de «un tercio de hombre» o «un tercio de flecha». Por otra parte, las magnitudes y los objetos continuos y homogéneos

son susceptibles de ser divididos y agrupados otra vez sin perder su carácter esencial. El puré de patatas ofrece un excelente ejemplo de objeto homogéneo que en sí mismo no está compuesto de partes, pero que puede, sin embargo, separarse en partes tan pequeñas como se desee. Las longitudes, las áreas y los volúmenes tienen la misma propiedad: aunque en esencia son continuos y no están realmente divididos en partes, ofrecen sin embargo la posibilidad de una división ilimitada.

Encontramos así dos clases contradictorias de objetos: por una parte, los objetos discretos, indivisibles; por otra, los objetos divisibles que, antes de ser separados en partes, son continuos. Naturalmente, estas características contradictorias van siempre unidas, puesto que no hay objetos completamente indivisibles ni completamente continuos. Sin embargo, dada la existencia real de tales características, sucede a menudo que una de ellas predomina sobre la otra.

El mero hecho de que la matemática separe las formas de sus contenidos determina ya una división de aquéllas en dos clases: las formas discretas y las formas continuas.

El modelo matemático de un objeto es la unidad, y el modelo matemático de una colección de objetos discretos es una suma de unidades, la imagen, podríamos decir, de lo puramente discreto, separada de todas las demás cualidades. Por otra parte, el modelo matemático original, fundamental, de la continuidad es la figura geométrica: la línea recta en el caso más sencillo.

Tenemos ante nosotros, por tanto, dos contrastes —lo discreto y lo continuo— y sus imágenes matemáticas abstractas: el número entero y la extensión geométrica. El proceso de medida consiste en la unión de estos dos contrastes: lo continuo es medido por unidades individuales. Pero estas unidades, que en principio son indivisibles, no bastan; es preciso introducir partes fraccionarias de la unidad original. De este modo surgen los números fraccionarios, y el concepto de número se desarrolla precisamente como resultado de la unión de los mencionados contrastes.

Luego, a un nivel más abstracto, aparece el concepto de intervalos inconmensurables y, como corolario de él, el número real, imagen abstracta del aumento ilimitado de la precisión utilizada al determinar una magnitud. Este concepto no se formó inmediatamente, y el largo camino de su desarrollo pasó por más de un conflicto entre estos dos elementos contradictorios, lo discreto y lo continuo.

En primer lugar, Demócrito concibió las figuras como formadas

por átomos y de este modo redujo lo continuo a lo discreto. Pero el descubrimiento de los intervalos inconmensurables condujo al abandono de tal representación. Después de este hallazgo dejó de pensarse en las magnitudes continuas como formadas por elementos individualizados, átomos o puntos, y tampoco fueron representadas por números, puesto que en aquel tiempo no se conocían más números que los enteros y los fraccionarios.

El contraste entre lo continuo y lo discreto volvió a aparecer en la matemática con fuerza renovada en el siglo XVII, cuando se colocaron los cimientos del cálculo diferencial e integral. Aquí fue lo infinitesimal lo que estaba en juego. En algunos casos lo infinitesimal era concebido como una partícula real, «realmente» infinitesimal, «indivisible», de la magnitud continua, algo así como los átomos de Demócrito, excepto que ahora el número de estas partículas se consideraba infinitamente grande. El cálculo de áreas y volúmenes, o en otras palabras, la integración, equivalía, pues, a la suma de un número infinito de estas partículas infinitamente pequeñas. Las áreas, por ejemplo, se entendían como «la suma de las líneas de las que está formada» (Fig. 3). Así, pues, se trataba de una nueva reducción de lo continuo a lo discreto, pero ahora de un modo más complicado,

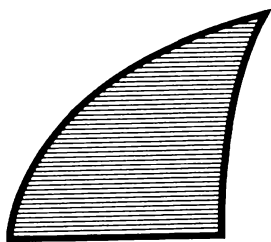


FIG. 3

a un nivel más alto. No obstante, este punto de vista también demostró ser insatisfactorio, y como contrapartida apareció (como consecuencia de los trabajos de Newton) la noción de variable *continua*, la noción de lo infinitesimal como variable *continua* que decrece sin límite. Esta concepción triunfó finalmente a comienzos del siglo XIX, época en que se fundó la teoría rigurosa de los límites. Los intervalos se concebían ahora como constituidos no de puntos o «indivisibles», sino como una extensión, como un medio continuo en el que sólo era posible aislar puntos

distintos o separar los diferentes valores de una magnitud continua. Los matemáticos hablaron entonces de la «extensión». En la unión de lo discreto y lo continuo, fue otra vez lo continuo lo que dominó.

Pero el desarrollo del análisis necesitaba una precisión mayor en la teoría de las magnitudes variables y sobre todo en la definición general de número real como valor posible de una magnitud variable. En los años setenta del siglo pasado surgió una teoría de los números reales que concebía la noción de intervalo como un conjunto de pun-

tos, y el rango de variación de una variable, como un conjunto de números reales. Lo continuo volvía a constar de puntos discretos, y las propiedades de la continuidad volvieron a ser expresadas por medio de la estructura del conjunto de puntos que lo formaban. Esta concepción condujo a un inmenso progreso en la matemática y llegó a adquirir un carácter dominante. Pero una vez más surgieron en ella profundas dificultades, y esto obligó a volver a la noción de continuidad pura, aunque ahora en un nivel distinto. Se hicieron diversos intentos de modificar el concepto de intervalo como conjunto de puntos y surgieron nuevos puntos de vista en torno a los conceptos de número, variable y función. Hoy día, el desarrollo de esta teoría sigue en marcha.

6. *Otros resultados de la interacción de la aritmética y la geometría*

La interacción de la geometría y de la aritmética sirvió para algo más que para formar el concepto de número real. Esta misma interacción de la geometría con la aritmética (o, más exactamente, con el álgebra) también influyó en la formación de los números negativos y de los complejos, esto es, de los números de la forma $a + b\sqrt{-1}$. Los números negativos se representan por puntos de una línea recta situados a la izquierda del punto que corresponde al cero. Fue exactamente esta representación geométrica la que proporcionó a los números imaginarios un puesto seguro en la matemática, pues hasta entonces no habían sido comprendidos. De aquella interacción surgieron nuevos conceptos como, por ejemplo, los vectores, que están representados por segmentos de recta orientados, y los tensores, cuya generalidad es aún mayor; en éstos se une otra vez el álgebra con la geometría.

La unión de varias teorías matemáticas ha desempeñado siempre un gran papel (en ocasiones decisivo) en el desarrollo de la matemática, y esto lo comprobaremos más tarde con la geometría analítica, el cálculo diferencial e integral, la teoría de funciones de variable compleja, el reciente análisis funcional, etc. Incluso en la misma teoría de números, esto es, en el estudio de los números enteros, se aplican con gran éxito métodos que dependen de la continuidad (por ejemplo, el análisis infinitesimal) y de la geometría. Estos métodos ocupan un lugar muy destacado en la teoría de números, en la «teoría analítica de números» y en la «geometría de números».

Desde un punto de vista ciertamente conocido, es posible contemplar los fundamentos de la matemática como la unión de conceptos

que surgen de la geometría y la aritmética; es decir, de los conceptos generales de continuidad, y de las operaciones algebraicas (como generalizaciones de las operaciones aritméticas). Pero no es éste el lugar para hablar de estas complicadas teorías. El propósito del presente capítulo ha sido el de dar una idea de la interacción general de los conceptos, de su unión y del conflicto entre nociones contradictorias en la matemática, tomando para ello como ejemplo la interacción de la aritmética y la geometría en el desarrollo del concepto de número.

§5. *La era de la matemática elemental*

1. *Los cuatro períodos de la matemática*

El desarrollo de la matemática no puede ser reducido a la simple acumulación de nuevos teoremas, sino que incluye cambios esencialmente cualitativos. Estos cambios cualitativos no llevan, sin embargo, a un proceso de destrucción o abolición de las teorías ya existentes, sino que confieren a éstas una profundidad y un grado de generalización mayores, estableciéndose teorías más generales, para las cuales los desarrollos precedentes han preparado ya el camino.

Adoptando un enfoque muy general, podemos distinguir en la historia de la matemática cuatro períodos fundamentales, cualitativamente distintos. Naturalmente, no es posible trazar fronteras exactas entre estas etapas, puesto que los rasgos esenciales de cada período aparecen más o menos gradualmente, pero la distinción entre las etapas y el paso de una a otra sí están del todo claros.

La primera etapa (o período) es la de la aparición de la matemática como ciencia teórica pura e independiente. Comienza en los tiempos más remotos y se extiende hasta el siglo V a. C., o quizá antes, cuando los griegos crearon las bases de una matemática «pura» con su conexión lógica entre teoremas y demostraciones (en ese siglo aparecieron, en particular, exposiciones sistemáticas de la geometría, como los «Elementos» de Hipócrates de Quío). Esta primera etapa fue el período de formación de la aritmética y la geometría, en la forma considerada anteriormente. En este tiempo la matemática consistía en una colección de reglas aisladas, deducidas de la experiencia, y directamente conectadas con la vida diaria. Estas reglas no formaban todavía un sistema lógicamente unificado, pues ese carácter teórico que tiene hoy la matemática gracias a las demostraciones

lógicas de sus teoremas, se formó muy lentamente a medida que se fue acumulando material para ello. La aritmética y la geometría no estaban separadas, sino íntimamente relacionadas la una con la otra.

El segundo período puede ser caracterizado como el de la matemática elemental, como el de la matemática de las magnitudes constantes; sus sencillos resultados fundamentales constituyen ahora el contenido de los cursos de bachillerato. Este período duró casi dos mil años y terminó en el siglo xvii con la aparición de la matemática «superior». Este es el período del que nos ocuparemos con detalle en esta sección, dedicando las siguientes al tercer y cuarto períodos, el del fundamento y desarrollo del análisis y el de la matemática contemporánea.

2. *La matemática en Grecia*

El período de la matemática elemental puede dividirse a su vez en dos partes, que se distinguen por su contenido básico: el del desarrollo de la geometría (hasta el siglo ii d. C.) y el del predominio del álgebra (desde el siglo ii al xvii). Respecto a las circunstancias históricas puede dividirse en tres partes, que pueden llamarse «griega», «oriental» y del «Renacimiento europeo». El período griego coincide en el tiempo con el florecimiento general de la cultura griega: empieza en el siglo vii a. C.; alcanza su culminación en el siglo iii antes de Cristo, en la época de los grandes geómetras de la Antigüedad (Euclides, Arquímedes y Apolonio), y termina en el siglo vi d. C. La matemática, y en especial la geometría, tuvo un extraordinario desarrollo en Grecia. Conocemos los nombres y los resultados de numerosos matemáticos griegos, aunque son pocos los trabajos genuinos que han llegado hasta nuestros días. Es de resaltar que Roma no aportó nada a la matemática, a pesar de que alcanzó su cenit en el siglo i d. C., cuando en Grecia —que había sido conquistada por Roma— florecía todavía la ciencia.

Los griegos no se limitaron a desarrollar y sistematizar la geometría elemental con una extensión como la de los «Elementos» de Euclides, que se enseñan hoy en día en los centros de bachillerato, sino que lograron, además, resultados de mucha más monta. Estudiaron las secciones cónicas: elipse, hipérbola y parábola; demostraron ciertos teoremas relativos a los elementos de lo que se llama geometría proyectiva; guiados por las necesidades de la astronomía, desarrollaron la geometría esférica (en el siglo i d. C.) y los elementos de la trigonometría, y calcularon las primeras tablas de senos (Hiparco en el

siglo II a. C. y Claudio Tolomeo en el siglo II d. C.)¹⁶; determinaron las áreas y volúmenes de diversas figuras complejas; por ejemplo, Arquímedes encontró el área del segmento de una parábola demostrando que es $2/3$ del área del rectángulo que lo contiene (Fig. 4). Los griegos también conocían el teorema que dice que, de todos los cuerpos de área superficial dada, la esfera es el de mayor volumen,

pero su demostración no se ha conservado ni era probablemente completa. Tal demostración es bastante difícil y se obtuvo por primera vez en el siglo XIX, gracias al cálculo integral.

En la aritmética y en el álgebra elemental, los griegos hicieron también considerables progresos. Como ya se ha mencionado, establecieron las bases de la teoría de números. A ésta pertenecen, por ejemplo, sus investigaciones sobre los números primos (el teorema de Euclides sobre la existencia de un número infinito de números primos y la «criba» de Eratóste-

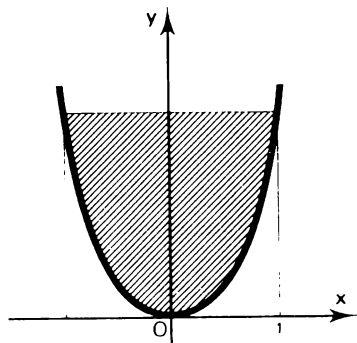


FIG. 4

nes para la obtención de los números primos) y la resolución de ecuaciones en números enteros (Diofanto, alrededor de 246-330 d. C.).

Ya hemos dicho que los griegos descubrieron las magnitudes irracionales, pero que sólo las consideraron geoméricamente, como segmentos de recta. Los problemas que hoy se abordan con ayuda del álgebra fueron tratados geoméricamente por los griegos. Y fue así como resolvieron las ecuaciones cuadráticas y transformaron las expresiones irracionales. Por ejemplo, la ecuación que hoy escribimos en la forma $x^2 + ax = b^2$, la establecieron como sigue: encontrar un segmento x tal que si al cuadrado construido sobre él se le suma un rectángulo construido sobre el mismo segmento y sobre un segmento dado a , obtengamos un rectángulo de área igual a la de un cuadrado dado. Este predominio de la geometría se conservó durante mucho tiempo, aun después de la época de los griegos. También estaban familiarizados con los métodos (geoméricos) de extracción de raíces

¹⁶ Tolomeo es bien conocido como autor de un sistema en el que la Tierra se considera el centro del universo y todos los cuerpos celestes se mueven alrededor de ella. Este sistema fue sustituido por el sistema de Copérnico.

cuadradas y cúbicas y con las propiedades de las progresiones aritméticas y geométricas.

Así, pues, los griegos estaban ya en posesión de gran parte del material del álgebra elemental contemporánea, pero no, en cambio, de los siguientes elementos esenciales: los números negativos y el cero, los números irracionales como concepto independiente de la geometría, y finalmente un sistema de símbolos literales bien desarrollado. Es cierto que Diofanto hizo uso de símbolos literales para expresar una cantidad desconocida y sus potencias, y también de símbolos especiales para la adición, substracción e igualdad, pero sus ecuaciones algebraicas estaban todavía escritas con coeficientes numéricos concretos.

En geometría los griegos alcanzaron el nivel que hoy denominamos matemática «superior». Arquímedes hizo uso del cálculo integral para el cálculo de áreas y volúmenes, y Apolonio utilizó la geometría analítica en sus investigaciones sobre secciones cónicas. Apolonio realmente da las ecuaciones de estas curvas¹⁷, pero las expresa en lenguaje geométrico. En estas ecuaciones no aparece todavía la noción general de constante arbitraria o de magnitud variable; y los medios necesarios para expresar tales conceptos, esto es, los símbolos literales del álgebra, no aparecen sino en una época posterior; sólo ellos podían convertir tales investigaciones en fuente de nuevas teorías, que pasarían ciertamente a formar parte de la matemática superior. Los fundadores de estas nuevas teorías se guiaron, mil años más tarde, por el legado de los científicos griegos; de hecho, la «Geometría» de Descartes (1637), fundador de la geometría analítica, empieza con una selección de problemas planteados por los griegos.

Tal es la regla general. Las viejas teorías, al dar lugar a nuevos y profundos problemas, se superan a sí mismas y exigen nuevas formas e ideas para seguir progresando. Pero estas formas e ideas pueden necesitar nuevas condiciones históricas para su nacimiento. En las sociedades antiguas las condiciones necesarias para pasar a la matemática superior ni existían ni podían existir; las circunstancias adecuadas surgieron con el desarrollo de las ciencias naturales en los tiempos modernos, desarrollo que, en los siglos XVI y XVII, estuvo condicionado a su vez por las nuevas demandas que impuso la tec-

¹⁷ Apolonio da las «ecuaciones» de las secciones cónicas referidas a un vértice. Por ejemplo, la «ecuación» de la parábola $y^2 = 2px$ se forma así: el cuadrado sobre el lado y es igual en área al rectángulo de lados $2p$ y x . Naturalmente, en lugar de los símbolos p , x , y utiliza los correspondientes segmentos.

nología y las manufacturas, y que vino de la mano del nacimiento y desarrollo del capitalismo.

Los griegos agotaron prácticamente todas las posibilidades de la matemática elemental, lo que explica el hecho de que el brillante progreso de la geometría se paralizara al comienzo de nuestra era y fuese reemplazado por la trigonometría y el álgebra en los trabajos de Tolomeo, Diofanto y otros. De hecho, los trabajos de Diofanto se pueden considerar como el comienzo del período en el cual el álgebra tuvo un papel preponderante. Pero la sociedad antigua, cercana ya a su declive, no pudo hacer que la ciencia avanzase gran cosa en esta dirección.

Debe hacerse notar que, pocos siglos antes, la aritmética ya había alcanzado un alto nivel en China. Los científicos chinos de los siglos II y I a. C. describieron las reglas para la resolución aritmética de un sistema de tres ecuaciones de primer grado. Es aquí donde por vez primera en la historia se hace uso de coeficientes negativos y se formulan reglas para operar con números negativos. Pero las soluciones se buscaban únicamente entre los números positivos, al igual que más tarde en los trabajos de Diofanto. Estos libros chinos incluyen también un método para la extracción de raíces cuadradas y cúbicas.

3. *El Oriente Medio*

Con el final de la ciencia griega empieza en Europa un período de estancamiento científico, desplazándose el centro del desarrollo matemático a la India, Asia central y los países árabes¹⁸. Durante un período de aproximadamente mil años —de los siglos V al XV— la matemática se desarrolló principalmente en conexión con las necesidades del cálculo, en particular de los cálculos astronómicos, puesto que los matemáticos orientales fueron en su mayor parte también astrónomos. Es cierto que no añadieron nada de importancia a la geometría griega; en este campo se limitaron a conservar y transmitir los resultados de los griegos. Pero los matemáticos indios, árabes y del Asia central lograron importantes éxitos en el campo de la aritmética y el álgebra¹⁹.

¹⁸ A efectos de orientación cronológica, damos aquí las épocas de algunos de los más relevantes matemáticos orientales. De la India: Aryabhata, nacido aproximadamente el 476 d. C.; Brahmagupta, aproximadamente 598-660; Bhaskara, siglo XII. De Kharizm: Al-Kharizmi, siglo IX; Al-Biruni, 973-1048. De Azerbaiján: Nasireddin Tusi, 1201-74. De Samarkand: Gyasseddin Jamschid, siglo XV.

¹⁹ Debe tenerse en cuenta que es un error asociar el desarrollo de la mate-

Como ya se mencionó en el §2, los indios inventaron nuestro sistema actual de numeración. Introdujeron también los números negativos, equiparando la diferencia entre los números negativos y positivos, la diferencia existente entre el haber y el debe o entre los dos sentidos de una línea recta. Finalmente, comenzaron a operar con magnitudes irracionales del mismo modo que con las racionales, sin representarlas geométricamente, a diferencia de los griegos. También tenían símbolos especiales para las operaciones algebraicas, incluyendo la extracción de raíces. El simple hecho de que los eruditos de la India y del Asia central no se vieran ya desconcertados por la diferencia entre magnitudes racionales e irracionales, les permitió superar ese «predominio» que había ejercido la geometría en la matemática griega, lo cual abrió nuevos caminos para el desarrollo del álgebra contemporánea.

El gran poeta y matemático Omar Khayyam (aproximadamente 1048-1122) y también el matemático azerbaijano Nasireddin Tusi (1201-1274), demostraron rigurosamente que todo cociente de magnitudes (sean conmensurables o inconmensurables) puede ser llamado número; en sus trabajos encontramos la misma definición de número, tanto para los racionales como para los irracionales, que la introducida anteriormente en el §4 en la formulación de Newton. La trascendencia de estos resultados se pone particularmente de manifiesto cuando recordamos que el descubrimiento completo de los números negativos e irracionales por parte de los matemáticos europeos fue un proceso sumamente lento, incluso después del comienzo del Renacimiento de la matemática en Europa. Por ejemplo, el célebre matemático francés Vieta (1540-1603), a quien el álgebra debe tanto, evitó sistemáticamente los números negativos, y en Inglaterra hubo una fuerte oposición a ellos hasta entrado el siglo XVIII. Estos números eran considerados absurdos puesto que eran menores que cero, esto es, «menores que la nada». En nuestros días constituyen ya un hecho familiar, aunque sólo sea en la forma de temperaturas negativas; todo el mundo, al leer los periódicos, entiende el significado de la expresión «la temperatura en Moscú es -8°C ».

La palabra «álgebra» proviene del nombre de un tratado del matemático y astrónomo Mahommed ibn Musa al-Kharizmi (Mahommed, hijo de Musa, nativo de Kharizm), que vivió en el siglo IX.

mática de este período principalmente con los árabes. El término «matemática arábica» se empezó a utilizar fundamentalmente porque la mayor parte de los sabios orientales escribían en lengua árabe, que se había extendido considerablemente tras las conquistas árabes.

Su tratado sobre álgebra llevaba por título *Al-jebr w'al-muqabala* que significa: «transposición y eliminación». Por transposición (*al-jebr*) se entiende la transferencia de términos al otro miembro de la ecuación, y por eliminación (*al-muqabala*) la cancelación de términos iguales en ambos miembros.

La palabra árabe *al-jebr* se convirtió en «álgebra» al transcribirla al latín, mientras que *al-muqabala* fue desechada, lo cual explica el término moderno «álgebra»²⁰ para esta disciplina.

El origen de este término responde muy bien al contenido real de la ciencia misma. El álgebra es en esencia la doctrina de las operaciones matemáticas consideradas formalmente desde un punto de vista general, con abstracción de los números concretos. Sus problemas están relacionados fundamentalmente con las reglas formales para la transformación de expresiones y la solución de ecuaciones. Al-Kharizmi colocó en la portada de su libro los nombres reales de dos de las reglas formales más generales, expresando de este modo el verdadero espíritu del álgebra.

Más tarde, Omar Khayyam definió el álgebra como la ciencia de resolver ecuaciones. Esta definición retuvo su significado hasta finales del siglo XIX, cuando el álgebra, junto con la teoría de ecuaciones, tomó nuevos derroteros, modificando esencialmente su carácter, pero no ese espíritu de generalidad que posee como ciencia de las operaciones formales.

Los matemáticos del Asia central encontraron métodos para el cálculo, tanto exacto como aproximado, de las raíces de las ecuaciones; descubrieron la fórmula general del «binomio de Newton», aunque la expresaron en palabras, y dieron un gran impulso y sistematizaron la trigonometría, calculando además con mucha precisión tablas de senos. Estas tablas fueron calculadas, por necesidades astronómicas, por el matemático Gyasseddin hacia 1427; Gyasseddin trabajó con el famoso astrónomo uzbekistano Ulug Begh e inventó también las fracciones decimales ciento cincuenta años antes de que fueran reinventadas en Europa.

Resumiendo: en el transcurso de la Edad Media se construyó casi por completo, en la India y en el Asia central, el sistema decimal de numeración (incluidas las fracciones), así como el álgebra elemental y la trigonometría. En el mismo período los resultados de la ciencia china empezaron a irrumpir en los países vecinos; hacia el siglo VI

²⁰ Conviene notar también que el término matemático «algoritmo», que indica un método o conjunto de reglas de cálculo, proviene asimismo del nombre de Al-Kharizmi.

antes de Cristo, los chinos ya conocían métodos de resolución de las ecuaciones indeterminadas más sencillas, métodos de cálculo aproximado en geometría y técnicas primitivas para la resolución aproximada de las ecuaciones de tercer grado. Las únicas partes de los actuales cursos de álgebra de bachillerato que no eran conocidas antes del siglo XVI son los logaritmos y los números imaginarios. Sin embargo, aún no existía un sistema de símbolos literales: el contenido del álgebra había dejado atrás su forma. Pero la forma era indispensable: la abstracción de los números concretos y la formulación de reglas generales necesitaban del correspondiente método de expresión; era esencial tener algún medio de denotar números *arbitrarios* y operaciones con ellos. El simbolismo algebraico es la forma adecuada al contenido del álgebra. Al igual que en la remota Antigüedad fue necesario inventar previamente símbolos para operar con los números enteros, así, para operar con números arbitrarios y dar reglas generales para su uso, fue necesario inventar los correspondientes símbolos. Esta tarea, iniciada en tiempos de los griegos, no se completó hasta el siglo XVII, cuando el sistema simbólico actual fue finalmente expuesto en los trabajos de Descartes y otros autores.

4. *La Europa del Renacimiento*

En tiempos del Renacimiento, los europeos tomaron contacto con la matemática griega a través de traducciones árabes. Los libros de Euclides, Tolomeo y Al-Kharizmi fueron traducidos en el siglo XII del árabe al latín —el lenguaje científico común en la Europa occidental— y, al mismo tiempo, el antiguo sistema de cálculo (derivado del de los griegos y los romanos) fue reemplazado gradualmente por el método indio, que fue tomado por los europeos de los árabes.

Fue sólo en el siglo XVI cuando la ciencia europea sobrepasó finalmente la de sus predecesores. Así, los italianos Tartaglia y Ferrari resolvieron la ecuación cúbica general y, más tarde, la ecuación general de cuarto grado (ver Cap. 4). Observemos que aunque estos resultados no se enseñan en la escuela, pertenecen, respecto a los métodos empleados en ellos, al álgebra elemental. Al álgebra superior debemos sin embargo referir la teoría general de las ecuaciones.

En el mismo período empezaron a utilizarse por primera vez los números imaginarios, lo que se hizo al principio de un modo puramente formal, sin fundamento lógico; este fundamento no llegaría sino a comienzos del siglo XIX. También se inventaron los símbolos algebraicos actuales; en particular, Vieta utilizó en el año 1591 sím-

bolos literales no sólo para las cantidades desconocidas, sino también para las dadas.

Muchos fueron los matemáticos que participaron en el desarrollo del álgebra. Al mismo tiempo aparecieron en Europa las fracciones decimales, inventadas por el sabio alemán Stevin, que escribió sobre ellas en 1585.

Finalmente Neper, en Gran Bretaña, inventó los logaritmos como ayuda para los cálculos astronómicos y escribió sobre ellos en 1614; Briggs calculó las primeras tablas de logaritmos decimales, que se publicaron en 1624²¹.

Al mismo tiempo aparecieron en Europa la «teoría de las combinaciones» y la fórmula general del «binomio de Newton»²²; toda vez que las progresiones ya se conocían, quedaba completada ya la estructura del álgebra elemental. Así tocó a su fin, a comienzos del siglo XVIII, todo el período de la matemática de las magnitudes constantes, de la matemática elemental tal como hoy día se enseña, con algunas adiciones, en nuestras escuelas. La aritmética, la geometría elemental, la trigonometría y el álgebra elemental quedaban así esencialmente completas. Después sobrevino la etapa de transición a la matemática superior, a la matemática de las magnitudes variables.

No se debe pensar, sin embargo, que el desarrollo de la matemática elemental cesó en esta época; por ejemplo, en la geometría elemental se descubrieron y se siguen descubriendo hoy día resultados nuevos. Además, gracias precisamente al subsiguiente desarrollo de la matemática superior, entendemos ahora más claramente la naturaleza esencial de la matemática elemental. Pero el papel principal en la matemática lo desempeñaban ahora los conceptos de magnitud variable, función y límite. Los problemas que llevaron de la matemática elemental a la matemática superior vienen aclarados y resueltos en nuestros días por los conceptos y métodos de la matemática superior (a veces son incluso irresolubles por los métodos ele-

²¹ Es interesante observar que Neper no definió los logaritmos como se hace actualmente cuando se dice que en la fórmula $x = a^y$ el número y es el logaritmo de x en la base a . Esta definición de logaritmo se dio más tarde. La definición de Neper estaba relacionada con los conceptos de magnitud variable e infinitésimo, y se expresaba diciendo que el logaritmo de x es una función $y = f(x)$ cuyo ritmo de crecimiento es inversamente proporcional a x ; esto es, $y' = c/x$ (ver capítulo 2). De este modo la base de la definición era esencialmente una ecuación diferencial, aunque las diferenciales todavía no se habían inventado.

²² La fórmula lleva el nombre de Newton no porque fuese éste el primero en descubrirla, sino porque la generalizó de exponentes enteros a exponentes fraccionarios e irracionales.

mentales), y hay otros problemas que pueden ser demostrados por los métodos de la matemática elemental, pero que incluso hoy sirven de punto de partida para resultados más generales e incluso para teorías enteras. Ejemplos de esto son la ya mencionada teoría de los sistemas regulares de las figuras y los problemas de la teoría de números que, a pesar de ser elementales en su formulación, están lejos de esta elementalidad en los métodos empleados en su resolución. Para más detalles remitimos al lector al capítulo 10.

§6. *La matemática de las magnitudes variables*

1. *Variable y función*

En el siglo xvi, el problema central de la física fue el estudio del movimiento. Las necesidades de la vida diaria y el desarrollo del conjunto de la ciencia condujeron a la física a este problema, así como a otros en los que aparece la interdependencia de magnitudes variables.

Como reflejo de las propiedades generales del concepto de cambio aparecen en la matemática los conceptos de magnitud variable y de función, y fue esta extensión capital del objeto de la matemática lo que determinó la transición a una nueva etapa: a la matemática de las magnitudes variables.

La ley del movimiento de un cuerpo en una trayectoria dada —por ejemplo, a lo largo de una línea recta— viene definida por el modo en que la distancia recorrida por el cuerpo aumenta con el tiempo.

Así, Galileo (1564-1642) descubrió la ley de la caída de los cuerpos, estableciendo que la distancia recorrida en la caída crece proporcionalmente al cuadrado del tiempo. Este hecho se expresa por la conocida fórmula

$$s = \frac{gt^2}{2}, \quad (1)$$

donde g es aproximadamente igual a $9,81 \text{ m/seg}^2$.

En general, la ley del movimiento expresa la distancia recorrida en el tiempo t . El tiempo t y la distancia s son, respectivamente, las variables «independiente» y «dependiente», y el hecho de que a cada instante t le corresponde una distancia definida s se expresa diciendo que la distancia s es función del tiempo t .

Los conceptos matemáticos de variable y función son generalizaciones abstractas de variables concretas (tales como tiempo, distancia, velocidad, ángulo de rotación y área de una superficie barrida)

y de las interdependencias entre ellas (la distancia depende del tiempo, etcétera). Igual que el concepto de número real es la imagen abstracta del valor real de una magnitud arbitraria, así una «variable» es la imagen abstracta de una magnitud que varía, lo que supone distintos valores durante el proceso en consideración. Una variable matemática x es «algo» o, más exactamente, «cualquier cosa» que puede tomar distintos valores numéricos. Este es el sentido general de variable; en particular podemos entender por ella el tiempo, la distancia o cualquier otra magnitud variable.

Exactamente del mismo modo, una función es la imagen abstracta de la dependencia de una magnitud respecto a otra. En la matemática, la afirmación de que y es función de x únicamente significa que a cada posible valor de x le corresponde un valor definido de y . Esta correspondencia entre los valores de y y los valores de x se llama función. Por ejemplo, de acuerdo con la ley de caída de los cuerpos, la distancia recorrida se corresponde con el tiempo de caída a través de la fórmula (1). La distancia es función del tiempo. Analicemos algunos otros ejemplos.

La energía de un cuerpo que cae se expresa por su masa y su velocidad de acuerdo con la fórmula

$$E = \frac{mv^2}{2}. \quad (2)$$

Para un cuerpo dado la energía es función de la velocidad v .

Según una conocida ley, la cantidad de calor generada en un conductor en la unidad de tiempo, al paso de una corriente eléctrica, se expresa por la fórmula

$$Q = \frac{RI^2}{2}, \quad (3)$$

donde I es la intensidad y R la resistencia del conductor. Para una resistencia dada, a cada intensidad de corriente I corresponde una cantidad definida de calor Q , generado en la unidad de tiempo. Esto es, Q es función de I .

El área de un triángulo rectángulo S con un ángulo agudo dado α y el lado correspondiente x (ver Fig. 5) viene expresada por la fórmula

$$S = \frac{1}{2} x^2 \operatorname{tg} \alpha. \quad (4)$$

Para un ángulo dado α , el área es función del lado x .

Todas estas fórmulas (1)-(4) se pueden resumir en una sola:

$$y = \frac{1}{2} ax^2. \quad (5)$$

Esta fórmula general representa el paso de las magnitudes variables concretas t, s, E, Q, v , etc. a las variables generales x e y , y de las relaciones concretas (1), (2), (3), (4) a la forma general (5). En Mecánica y Electricidad se utilizan las fórmulas concretas (1), (2), (3), que relacionan magnitudes concretas, mientras que la teoría matemática de las funciones opera con la fórmula general (5), sin asociar esta fórmula a ninguna magnitud determinada.

El siguiente grado de abstracción de lo concreto consiste en examinar, no una dependencia dada de la y respecto a la x , como $y = \frac{1}{2} ax^2$, $y = \sin x$, $y = \log x$, etc., sino la correspondencia general de y con x , expresada en la fórmula abstracta

$$y = f(x).$$

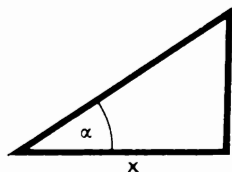


FIG. 5

Esta fórmula establece que la magnitud y es, de modo general, función de x ; esto es, cada valor dado de la x se corresponde, de una manera o de otra, con un valor definido de la y . Lo importante en la matemática no son ciertas funciones particulares ($y = \frac{1}{2} ax^2$, $y = \sin x$, etc.), sino las funciones *en general*. Estos grados de abstracción —primero de las magnitudes concretas y luego de las funciones concretas— son análogos a los que se observan en la formación del concepto de número entero: primero, la abstracción de las colecciones concretas de objetos conduce al concepto de número entero (1, 3, 12, etc.); y luego, una abstracción posterior conduce al concepto de número entero en general. Esta generalización es el resultado de una profunda interacción entre análisis y síntesis: análisis de las relaciones mutuas, y síntesis, en forma de nuevos conceptos, de sus rasgos comunes.

La rama de la matemática dedicada al estudio de las funciones se llama análisis, o más comúnmente análisis infinitesimal, puesto que uno de los elementos más importantes en el estudio de las funciones es el concepto de lo infinitesimal (el sentido de este concepto y su significado se explicarán en el capítulo 2).

Puesto que una función es la imagen abstracta de la dependencia de una magnitud respecto a otra, podemos decir que el análisis tiene como objeto de estudio las dependencias entre magnitudes variables, no entre una magnitud concreta y otra, sino entre variables en general, con abstracción de su contenido. Una abstracción de esta clase garantiza gran amplitud en sus aplicaciones, puesto que una fórmula o un teorema contienen un número infinito de casos concretos posibles. Ya se ha dado un ejemplo de esto en las sencillas fórmulas (1)-(5). De este modo resulta evidente la completa analogía entre el análisis, por una parte, y la aritmética y el álgebra, por otra. Todas estas ramas se originaron a partir de problemas concretos y proporcionan expresiones abstractas generales de las relaciones concretas del mundo real.

2. Geometría analítica y análisis

El nuevo período de la matemática, que comienza en el siglo xvii, puede ser definido como el período del nacimiento y desarrollo del análisis. (Este es el tercero de los tres períodos importantes antes mencionados.) Se sobreentiende, naturalmente, que ninguna teoría surge como resultado de la simple formulación de nuevos conceptos y que el análisis no podría resultar de la mera existencia de los conceptos de variable y función. La creación y desarrollo de una teoría —y tanto más de una rama completa de la ciencia, como la del análisis matemático— requieren que los nuevos conceptos cobren una actividad propia, es decir, que entre ellos se descubran nuevas relaciones que permitan la solución de nuevos problemas.

Es más: un nuevo concepto sólo puede nacer, desarrollarse y ganar en generalidad y precisión sobre la base de esos mismos problemas que aquél permite resolver, y sólo por medio de los teoremas de los que forma parte. Los conceptos de variable y función no surgieron en su forma definitiva ni en la mente de Galileo, Descartes o Newton, ni en la de cualquier otro matemático concreto. Los intuieron muchos matemáticos (por ejemplo, Neper en conexión con los logaritmos) y gradualmente tomaron una forma más o menos definida, aunque no definitiva, en Newton y Leibnitz, haciéndose aún más precisos y generales en el subsiguiente desarrollo del análisis. La definición actual data sólo del siglo xix, pero ni es *totalmente* rigurosa ni seguramente la última. El desarrollo del concepto de función continúa incluso en el momento actual.

El análisis matemático se basó en los materiales suministrados por la nueva ciencia de la mecánica y en problemas de la geometría y el álgebra. El primer paso definido hacia la matemática de las

magnitudes variables fue la aparición, en 1637, de la «Geometría» de Descartes, donde se establecían las bases de la llamada geometría analítica. Las ideas básicas de Descartes son las siguientes.

Supongamos que nos dan, por ejemplo, la ecuación

$$x^2 + y^2 = a^2. \quad (6)$$

En álgebra la x y la y se consideraban como incógnitas; y puesto que la ecuación no permite determinarlas, ésta no ofrecía ningún interés especial para el álgebra. Pero Descartes no consideró x e y como incógnitas a obtener de la ecuación inicial, sino como *variables*; de este modo la anterior ecuación expresa, en este sentido, la interdependencia de dos variables. Tal ecuación puede escribirse en forma general, pasando todos sus términos al primer miembro:

$$F(x, y) = 0.$$

Además, Descartes introdujo en el plano las coordenadas x, y , que ahora llamamos cartesianas (Fig. 6). De este modo, a cada par de valores x e y corresponde un punto, y recíprocamente, a cada punto corresponde un par de coordenadas x, y . Así, la ecuación $F(x, y) = 0$ determina el lugar geométrico de los puntos del plano cuyas coordenadas satisfacen la ecuación. En general, éste será una curva. Por ejemplo, la ecuación (6) determina la circunferencia de radio a y centro el origen. En efecto, como se ve en la figura 7, $x^2 + y^2$ es, por el teorema de Pitágoras, el cuadrado de la distancia del origen O al punto M de coordenadas x e y . Así la ecuación (6) representa el lugar geométrico de los puntos cuya distancia al origen es igual a a , y este lugar es una circunferencia.

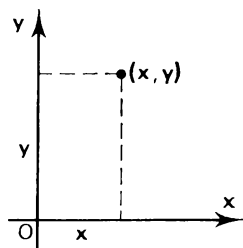


FIG. 6

Recíprocamente, un lugar geométrico que venga dado por una condición geométrica puede definirse también mediante una ecuación que exprese la misma condición en lenguaje algebraico, por medio de coordenadas. Por ejemplo, la condición geométrica que define una circunferencia como lugar geométrico de los puntos equidistantes de un punto dado, puede ser expresada, en lenguaje algebraico, por la ecuación (6).

Así, el problema y el método general de la geometría analítica son los siguientes: representar una ecuación dada de dos variables por una curva en el plano y deducir, de las propiedades algebraicas de la ecuación, las propiedades geométricas de la curva correspondiente; y recíprocamente, de las propiedades geométricas de la curva obtener la ecuación, y entonces, de las propiedades algebraicas de la ecuación, deducir las propiedades geométricas de la curva. De este modo, los

problemas geométricos pueden reducirse a problemas algebraicos y, en definitiva, a cálculo.

El contenido de la geometría analítica se discutirá con detalle en el capítulo 3. Llamemos ahora la atención acerca del hecho de que aquella (como se deduce de la breve explicación anterior) se originó por la unión de la geometría, el álgebra y la idea de magnitud variable. El principal contenido geométrico de los comienzos de la geometría

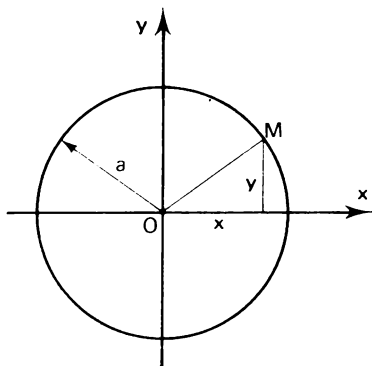


FIG. 7

analítica fue la teoría de las secciones cónicas: elipse, hipérbola y parábola. Esta teoría, como ya hemos mencionado, fue desarrollada por los griegos; los resultados de Apolonio ya contenían en forma geométrica las ecuaciones de las secciones cónicas. La unión de este contenido geométrico con la forma algebraica, desarrollada después de la época griega, y con la idea de magnitud variable, que surgió del estudio del movimiento, originó la geometría analítica.

Entre los griegos las secciones cónicas eran objeto de un interés puramente matemático; pero en la época de Descartes tenían ya una importancia práctica para la astronomía, la mecánica y la tecnología. Kepler (1571-1630) descubrió que los planetas se mueven alrededor del Sol en elipses, y Galileo estableció el hecho de que un cuerpo lanzado al aire —lo mismo si es una piedra que una bala de cañón— se mueve según una parábola (en primera aproximación, despreciando la resistencia del aire). Como resultado de todo ello, se convirtió en urgente necesidad el cálculo de diversos parámetros referentes a las secciones cónicas, y fue el método de Descartes el que resolvió este

problema. Así, pues, el desarrollo precedente de la matemática preparó el camino para su método; y éste, por su parte, surgió de las insistentes demandas de la ciencia y la tecnología.

3. *Cálculo diferencial e integral*

El siguiente paso decisivo en la matemática de las magnitudes variables fue dado por Newton y Leibnitz durante la segunda mitad del siglo XVII al sentar las bases del cálculo diferencial e integral. Este fue el verdadero comienzo del análisis, puesto que el objeto de este cálculo son las propiedades de las funciones mismas, distinto del objeto de la geometría analítica, que son las figuras geométricas. De hecho, lo que hicieron Newton y Leibnitz fue completar esa cantidad inmensa de trabajo que habían desarrollado hasta entonces muchos matemáticos y que se extendía hasta los métodos de determinación de áreas y volúmenes empleados por los antiguos griegos.

No desarrollaremos ahora los conceptos fundamentales del cálculo diferencial e integral ni las teorías del análisis que se derivaron de ellos, puesto que esto se hará en los capítulos dedicados a estas teorías. Aquí sólo queremos llamar la atención acerca de los orígenes de este cálculo, que fueron principalmente los nuevos problemas de la mecánica y los viejos problemas de la geometría, consistentes estos últimos en la determinación de tangentes a una curva dada y en el cálculo de áreas y volúmenes. Estos problemas geométricos habían sido ya estudiados por los antiguos (basta mencionar a Arquímedes), y también por Kepler, Cavalieri y otros, a principios del siglo XVII. Pero el factor decisivo fue el descubrimiento de una notable relación entre estos dos tipos de problemas y la formulación de un método general para resolverlos; tal fue la obra de Newton y Leibnitz.

Esta relación, que permitió conectar los problemas de la mecánica con los de la geometría, fue descubierta gracias a la posibilidad (brindada por el método de las coordenadas) de hacer una representación gráfica de la dependencia de una variable respecto de otra o, en otras palabras, de una función. Con la ayuda de esta representación gráfica es fácil formular la relación antes mencionada entre los problemas de la mecánica y la geometría (relación que fue el origen del cálculo diferencial e integral) y describir así el contenido general de estos dos tipos de cálculo.

El cálculo diferencial es, básicamente, un método para encontrar la velocidad de un movimiento cuando se conoce la distancia recorrida

en un tiempo dado. Este problema se resuelve por «derivación» y es completamente equivalente al problema de dibujar una tangente a la curva que representa la dependencia de la distancia respecto del tiempo. La velocidad en el instante t es igual a la pendiente de la tangente a la curva en el punto correspondiente a t (Fig. 8).

El cálculo integral es en esencia un método para encontrar la distancia recorrida cuando se conoce la velocidad y, en general, de encontrar el resultado total de la acción de una magnitud variable. Evidentemente, este problema es el recíproco del problema del

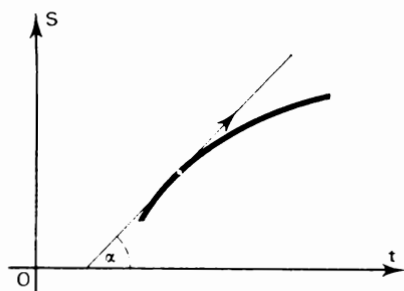


FIG. 8

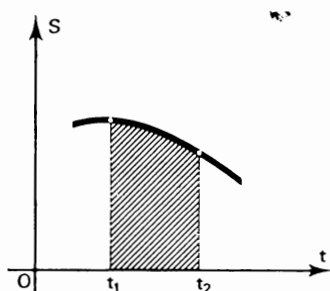


FIG. 9

cálculo diferencial (el problema de encontrar la velocidad), y se resuelve por «integración». Resulta que el problema de la integración es en todo equivalente al de encontrar el área bajo la curva que representa la dependencia de la velocidad respecto del tiempo. La distancia recorrida en el intervalo del tiempo t_1 a t_2 es igual al área bajo la curva entre las rectas que corresponden en la gráfica a los valores t_1 y t_2 (Fig. 9).

Haciendo abstracción de la formulación mecánica de los problemas y operando con funciones en vez de dependencias de distancia o velocidad respecto al tiempo, se obtienen los problemas del cálculo diferencial e integral en forma abstracta.

Fundamental para el cálculo, como para todo el desarrollo posterior del análisis, es el concepto de límite, que fue formulado algo más tarde que los otros conceptos fundamentales de variable y función. En los primeros días del análisis, el papel que más tarde desempeñaría el límite, corrió a cargo de ese concepto algo nebuloso que es

el infinitésimo. Los métodos para el cálculo real de la velocidad, conocida la distancia recorrida (a saber, la derivación), y de la distancia, conocida la velocidad (integración), se basaban en la unión del álgebra con el concepto de límite. El análisis se originó por la aplicación de estos conceptos y métodos a los referidos problemas de la mecánica y la geometría (y también a otros problemas: por ejemplo, los de máximos y mínimos). El análisis fue a su vez absolutamente necesario para el desarrollo de la mecánica, en la formulación de cuyas leyes ya se encontraban los conceptos analíticos en forma latente. Por ejemplo, la segunda ley de Newton, tal como él la formuló, establece que «la variación de la cantidad de movimiento es proporcional a la fuerza actuante» (con más precisión: el ritmo de variación del impulso es proporcional a la fuerza). Por consiguiente, si deseamos hacer uso de esta ley debemos estar en condiciones de definir el ritmo de variación de una variable, esto es, de derivarla. (Si establecemos la ley diciendo que la aceleración es proporcional a la fuerza, el problema es el mismo, porque la aceleración es proporcional al ritmo de variación del impulso.) También está perfectamente claro que, para establecer la ley que rige un movimiento cuando la fuerza es variable (en otras palabras, cuando el movimiento tiene lugar con aceleración variable), es preciso resolver el problema inverso de encontrar una magnitud dado su ritmo de variación; en otras palabras, es preciso integrar. Así, pues, se puede decir que Newton se vio simplemente *obligado* a inventar la derivación y la integración con el fin de poder desarrollar la mecánica.

4. *Otras ramas del análisis*

Junto con el cálculo diferencial e integral, surgieron otras ramas del análisis: la teoría de las series (ver Cap. 2, §14), la teoría de las ecuaciones diferenciales (Caps. 5 y 6), y la aplicación del análisis a la geometría, que se convirtió más tarde en una rama especial de la geometría, la geometría diferencial, que trata de la teoría general de las curvas y superficies (Cap. 7). Todas estas teorías surgieron para resolver problemas planteados por la mecánica, la física y la tecnología.

La teoría de las ecuaciones diferenciales —la rama más importante del análisis— estudia aquellas ecuaciones en las que la incógnita no es un número sino una función o, en otras palabras, una ley que expresa la dependencia de una magnitud respecto de otra o de varias otras. Es fácil comprender cómo surgen tales ecuaciones. En mecánica se trata de determinar la ley del movimiento de un cuerpo bajo condi-

ciones dadas, y no sólo valores particulares de la velocidad o de la distancia recorrida. En mecánica de fluidos es necesario encontrar la distribución de velocidades en toda la masa del fluido en movimiento; en otras palabras, encontrar la dependencia de la velocidad respecto a las tres coordenadas espaciales y del tiempo. Análogamente, en la teoría de la electricidad y el magnetismo se trata de encontrar el potencial en todos los puntos del espacio; esto es, la dependencia de este potencial respecto a las tres coordenadas espaciales.

Problemas de esta clase surgen continuamente en las distintas ramas de la mecánica (incluidas la hidrodinámica y la teoría de la elasticidad), en acústica, en la teoría de la electricidad y el magnetismo y en la del calor. Desde el mismo momento de su nacimiento, el análisis permaneció en íntimo contacto con la mecánica y con la física en general, y sus resultados más importantes estuvieron siempre relacionados con la solución de los problemas propuestos por las ciencias físicas. Empezando con Newton, los más grandes analistas, D. Bernoulli (1700-1782), L. Euler (1707-1783), J. Lagrange (1736-1813), H. Poincaré (1854-1912), M. V. Ostrogradski (1801-1861), y A. M. Lyapunov (1857-1918), así como muchos otros que obtuvieron nuevos resultados en análisis, empezaron, por regla general, por los problemas que tenía planteados la física contemporánea.

De este modo surgieron nuevas teorías: en conexión con la mecánica, Euler y Lagrange fundaron una nueva rama del análisis llamada cálculo de variaciones (ver Cap. 8), y al final del siglo XIX, Poincaré y Lyapunov, partiendo otra vez de problemas mecánicos, fundaron la llamada teoría cualitativa de las ecuaciones diferenciales (ver Cap. 5, §7).

En el siglo XIX el análisis se enriqueció con una rama muy importante, la teoría de las funciones de una variable compleja (ver Cap. 9). Rudimentos de ella se encuentran en los trabajos de Euler y otros matemáticos, pero su transformación en una teoría coherente tuvo lugar a mediados del siglo XIX, labor que corrió a cargo sobre todo del matemático francés Cauchy (1789-1857). Esta teoría no tardó en experimentar un desarrollo importante, proporcionando resultados que permitieron a los matemáticos penetrar más profundamente en muchas de las leyes del análisis, encontrando importantes aplicaciones a problemas de la matemática, de la física y la tecnología.

El análisis se desarrolló rápidamente; no sólo se convirtió en el centro y en la parte más importante de la matemática, sino que también irrumpió en regiones ya establecidas: el álgebra, la geometría e incluso la teoría de números. El álgebra empezó a ser considerada

como la teoría de las funciones expresadas en forma de polinomios de una o varias variables²³. La geometría analítica y diferencial comenzó a dominar el campo de la geometría. Desde los tiempos de Euler, los métodos del análisis se fueron infiltrando en la teoría de números, formando así la base de la llamada teoría analítica de números, que contiene algunos de los resultados más profundos de la ciencia de los números enteros.

Influenciada por el análisis, con sus conceptos de variable, función y límite, toda la matemática quedó impregnada de la idea de movimiento y cambio, es decir: de dialéctica. Del mismo modo, y fundamentalmente a través del análisis, la matemática se vio influenciada por las ciencias físicas y la tecnología a la par que participó en el desarrollo de éstas, puesto que fue el medio de expresar exactamente sus leyes y de resolver sus problemas. Igual que entre los griegos la matemática era fundamentalmente geometría, se puede decir que después de Newton fue, sobre todo, análisis. Naturalmente, el análisis no absorbió por completo a toda la matemática; en la geometría, en la teoría de números y en el álgebra subsistieron los problemas y métodos característicos de estas ciencias. Así, en el siglo XVII surgió junto a la geometría analítica otra rama de la geometría, la llamada geometría proyectiva, en la que predominaban los métodos puramente geométricos. La geometría proyectiva surgió en esencia de problemas de representación de objetos sobre un plano (proyección), por lo cual es particularmente útil en la geometría descriptiva.

Al mismo tiempo se desarrolló una nueva e importante rama de la matemática, la teoría de la probabilidad, que tiene como objeto las regularidades que se observan en grandes masas de fenómenos, tales como una serie de disparos de rifle o de lanzamientos de una moneda al aire. En el período subsiguiente adquiere especial importancia en la física y en la tecnología, y su desarrollo estuvo condicionado por los problemas que le planteaban otras ramas de la ciencia. El rasgo característico de esta teoría es que estudia las leyes de los «sucesos aleatorios», suministrando métodos matemáticos para la investiga-

²³ Los polinomios son funciones de la forma $y = a_0x^n + a_1x^{n-1} + \dots + a_n$. El problema fundamental del álgebra de entonces (la solución de la ecuación $a_0x^n + a_1x^{n-1} + \dots + a_n = 0$) significaba sencillamente buscar los valores de x para los cuales la función $y = a_0x^n + a_1x^{n-1} + \dots + a_n$ fuese igual a cero. La existencia de dicha solución, de una raíz de la ecuación, viene establecida en el teorema fundamental del álgebra, que se demuestra por medio del análisis (ver capítulo 4, §3).

ción de las irregularidades que necesariamente aparecen en sucesos de aquel tipo. Los rasgos básicos de la teoría de las probabilidades se explicarán en el capítulo 11.

5. *Aplicaciones del análisis*

El análisis, con sus distintas ramas, proporcionó a la física y a la tecnología métodos poderosos para la solución de problemas de muchas clases diferentes. Ya hemos mencionado los primeros: encontrar el ritmo de variación de una magnitud cuando se conoce su dependencia respecto del tiempo; encontrar el área de las figuras curvilíneas y los volúmenes de los sólidos, y hallar el resultado total de un proceso o la acción total de una magnitud variable. Así, el cálculo integral permite determinar el trabajo realizado por un gas en expansión cuando la presión varía según una ley conocida, o bien calcular, por ejemplo, el potencial de un campo eléctrico producido por un sistema dado de cargas, a partir de la ley de Coulomb (que da el potencial en un campo creado por una carga puntual), etc.

Además, el análisis suministra un método para encontrar los valores máximos y mínimos de una variable bajo condiciones dadas. Así, con la ayuda del análisis es fácil determinar las dimensiones de una cisterna cilíndrica que para un volumen dado tenga una superficie mínima y que requiera, por tanto, menos cantidad de material. Se deduce que la cisterna tendrá esta propiedad si su altura es igual al diámetro de su base (Fig. 10). El análisis permite también determinar la forma de la curva para la cual un cuerpo que se desliza de un punto dado a otro tarde el menor tiempo posible (esta curva es la llamada cicloide; Fig. 11).

Para la solución de estos y otros problemas remitimos al lector a los capítulos 2 y 8.

El análisis, o con más precisión, la teoría de las ecuaciones diferenciales, no sólo permite encontrar valores concretos de magnitudes variables, sino también determinar funciones desconocidas; esto es, encontrar leyes de dependencia de ciertas variables respecto de otras. Así tenemos la posibilidad, de acuerdo con las leyes generales de la electricidad, de calcular la variación con el tiempo de la corriente en un circuito con resistencia, capacidad y autoinducción dadas; podemos determinar las leyes de distribución de velocidades en toda la masa de un fluido en condiciones dadas; podemos deducir leyes generales para la vibración de cuerdas y membranas y para la propagación de vibraciones en los distintos medios (nos referimos a las ondas sonoras, a las electromagnéticas y a las vibraciones elásticas propaga-

das por la Tierra y producidas por terremotos o explosiones). Entre paréntesis, señalemos que de esta forma se obtienen nuevos métodos aplicables a la búsqueda de minerales útiles y al estudio del interior de la Tierra. En los capítulos 5 y 11 se encontrarán problemas de esta clase.

Finalmente, el análisis no sólo suministra métodos para resolver problemas específicos; también da métodos generales para la formulación matemática de las leyes cuantitativas de las ciencias físicas.

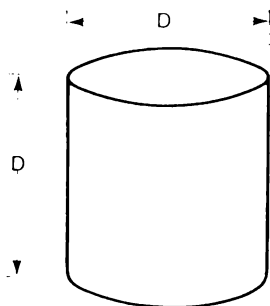


FIG. 10

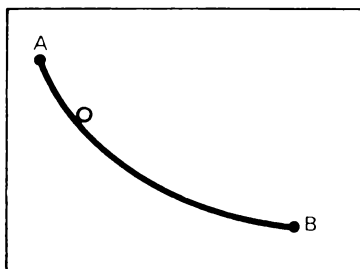


FIG. 11

Como ya se mencionó, las leyes generales de la mecánica no podrían formularse matemáticamente sin recurrir a los conceptos del análisis, y sin tal formulación no sería posible resolver los problemas de la mecánica. Del mismo modo, tampoco las leyes generales de la conducción del calor, de la difusión a través de materiales porosos, de la propagación de vibraciones, del curso de las reacciones químicas, de las leyes básicas del electromagnetismo ni otras muchas leyes podrían formularse matemáticamente sin los conceptos del análisis. Sólo como resultado de esta formulación pueden aplicarse dichas leyes a una gran variedad de casos concretos, suministrando la base para conclusiones matemáticamente precisas en problemas específicos como los de conducción del calor, vibraciones, soluciones químicas, campos electromagnéticos y otros problemas de la mecánica, astronomía y todas las numerosas ramas de la física, química, ingeniería del calor, energía, construcción de máquinas, ingeniería eléctrica, etc.

6. Examen crítico de los fundamentos del análisis

Igual que en la historia de la geometría griega la presentación rigurosa y sistemática dada por Euclides completó un largo desarrollo previo, así en la evolución del análisis surgió la necesidad de situar

a esta disciplina sobre una base más firme que la que proporcionaron los creadores de esta poderosa herramienta: Newton, Euler, Lagrange y otros. Cuando el análisis, fundado por ellos, empezó a adquirir cierto auge, comenzó, por una parte, a ocuparse de problemas más profundos y difíciles, y por otra, a necesitar, por su mayor extensión, de una base más sistemática y cuidadosamente razonada. El desarrollo de la teoría precisaba de una sistematización y un análisis crítico de sus fundamentos. Situar una teoría sobre una base firme requiere el examen de todo su desarrollo, y dicha base no debería ser considerada de modo alguno como punto de partida para la teoría misma, puesto que, si ésta no existiera, no sabríamos qué es lo que necesita ser fundamentado. De paso diremos que ciertos formalistas contemporáneos olvidan este hecho cuando consideran aconsejable fundar y desarrollar una teoría partiendo de axiomas que no se han seleccionado sobre la base de un análisis del material real que ellos pretenden sistematizar. Los axiomas mismos requieren una justificación de su contenido; ellos no hacen sino sintetizar el material disponible y suministrar la base para la construcción lógica de una teoría ²⁴.

Este período imprescindible de crítica, sistematización y fundamentación tuvo lugar, en el caso del análisis, a comienzos del siglo pasado. Gracias al esfuerzo de gran número de eminentes científicos, esta difícil e importante labor se realizó con completo éxito. En particular se dieron definiciones precisas de los conceptos básicos de número real, variable, función, límite y continuidad.

Sin embargo, como ya tuvimos ocasión de mencionar, ninguna de estas definiciones puede ser considerada como absolutamente rigurosa o definitiva. El desarrollo de estos conceptos continúa en la actualidad. Euclides, y todos los matemáticos que vivieron en los dos mil años siguientes, consideraron sin duda sus «Elementos» como el límite práctico del rigor lógico. Pero desde el punto de vista contemporáneo los fundamentos euclidianos de la geometría parecen bastante superficiales. Este ejemplo histórico nos demuestra que no debemos dejarnos seducir por la idea de que la matemática contemporánea posee un rigor «absoluto». Una ciencia que no está todavía muerta y momificada no es, ni puede ser de forma alguna, perfecta. Pero podemos decir con confianza que, primero, los fundamentos del análisis tal como son actualmente se corresponden de modo bastante

²⁴ Este doble papel de los axiomas se pierde a veces de vista incluso en trabajos de carácter metodológico, los cuales atribuyen así a los axiomas un significado que no les pertenece, a saber, el de la construcción total de una teoría.

satisfactorio con los problemas contemporáneos de la ciencia y con la concepción moderna de la precisión lógica; y segundo, que la continuada profundización en estos conceptos y las discusiones que hoy día están teniendo lugar en torno a ellos no llevan ni llevarán, sencillamente, a desecharlos; estas discusiones conducirán a un nuevo conocimiento, más preciso y profundo, cuyo resultado es todavía difícil de prever.

Aunque el establecimiento de los principios básicos de una teoría constituye un resumen de su desarrollo, no representa su final; por el contrario, conduce a un ulterior desarrollo. Esto es exactamente lo que sucedió en el análisis. En conexión con la profundización de sus fundamentos surgió una nueva teoría matemática, creada por el matemático alemán Cantor en los años setenta del siglo pasado: se trata de la teoría general de los conjuntos infinitos de objetos abstractos, sean números, puntos, funciones o cualesquiera otros «elementos». Sobre la base de estas ideas se abrió un nuevo capítulo del análisis, el de la llamada teoría de las funciones de variable real, cuyos conceptos, junto con los fundamentos del análisis y la teoría de conjuntos, se analizarán en el capítulo 14. Las ideas generales de la teoría de conjuntos penetraron en todas las ramas de la matemática. Pero este «punto de vista conjuntista» va inseparablemente unido a una nueva etapa en el desarrollo de la matemática, que consideraremos brevemente a continuación.

§7. *La matemática contemporánea*

1. *El carácter avanzado de la matemática actual*

Las cuatro etapas del desarrollo de la matemática mencionadas en el §5 se corresponden naturalmente con las distintas etapas de nuestra formación matemática; el material examinado en cada una de las etapas de nuestro estudio refleja con bastante aproximación el contenido básico del período correspondiente de la historia de la matemática.

Los resultados básicos de la aritmética y la geometría, obtenidos en el primer período del desarrollo de la matemática, constituyen el objeto de la enseñanza primaria, y son de todos conocidos. Por ejemplo, cuando determinamos la cantidad de material necesario para realizar un cierto trabajo (para recubrir un suelo, pongamos por caso) ya estamos haciendo uso de estos primeros resultados de la matemática. Los más importantes logros del segundo período,

el de la matemática elemental, se estudian en los centros de enseñanza media. Los resultados básicos del tercer período, los fundamentos del análisis, la teoría de las ecuaciones diferenciales, el álgebra superior, etcétera, constituyen los conocimientos matemáticos de un ingeniero; se estudian en todos los centros de enseñanza superior, excepto aquellos dedicados exclusivamente a las humanidades. De este modo, todo ingeniero y científico conoce y emplea las ideas básicas y los logros de la matemática de ese período.

Por otra parte, las ideas y los resultados del período actual de la matemática se estudian casi exclusivamente en los departamentos universitarios de matemáticas y física. Aparte de los especialistas en matemáticas, dichos resultados son utilizados por los investigadores en los campos de la mecánica y la física, así como en algunas de las ramas más recientes de la tecnología. Naturalmente, esto no significa en absoluto que no tengan aplicación práctica, sino que, como representan los resultados más recientes de la ciencia, poseen un carácter más complicado. Por consiguiente, al pasar ahora a una descripción general de la última etapa del desarrollo de la matemática no podremos esperar que todo lo que mencionemos en pocas palabras resulte del todo claro. Intentaremos describir, en pocas líneas y de modo sumamente general, el carácter de las nuevas ramas de la matemática; su contenido se explicará con mayor detalle en los correspondientes capítulos del libro.

Si el lector encuentra excesivamente difícil la lectura de esta sección, puede prescindir de ella por ahora y volver más tarde, una vez estudiados los capítulos pertinentes.

2. *Geometría*

El comienzo del desarrollo actual de la matemática está caracterizado por profundos cambios de todos sus campos: álgebra, geometría y análisis. Este cambio se observa claramente en el campo de la geometría. En el año 1826, Lobachevski —y casi simultáneamente con él el matemático húngaro Janos Bolyai— desarrolló la nueva geometría no-euclídea. Las ideas de Lobachevski estuvieron lejos de ser entendidas inmediatamente por todos los matemáticos. Eran demasiado atrevidas e inesperadas. Pero desde ese momento la geometría avanzó por unos cauces fundamentalmente nuevos, incluso cambió el concepto mismo de lo que se entendía por geometría. Su objeto y el campo de sus aplicaciones se extendieron rápidamente. El paso más importante en esta dirección, después de Lobachevski, lo dio en 1854 el famoso matemático alemán Riemann,

formulando con toda precisión la idea general de que la geometría permitía estudiar un número ilimitado de «espacios» e indicando al mismo tiempo sus posibles significados en el mundo real. El nuevo desarrollo de la geometría tenía dos rasgos característicos.

En primer lugar, la geometría primitiva estudiaba sólo las formas *espaciales* y las relaciones del mundo material, y ello únicamente en la medida en que aparecían en el marco de la geometría euclídea. Pero, después de Riemann, la geometría comenzó a incluir también *otras* muchas formas y relaciones del mundo real, sometidas a la única condición de que fuesen semejantes a las espaciales y permitiesen, por tanto, la utilización de los métodos geométricos. El término «espacio» adquirió así en la matemática un nuevo sentido más amplio y al mismo tiempo más específico. Simultáneamente, los métodos de la geometría se hicieron más ricos y variados, proporcionando a su vez medios más completos para comprender el mundo que nos rodea, el mundo del que fue abstraída la geometría en su forma original.

En segundo lugar, se hicieron importantes progresos incluso en la geometría euclídea, en cuyo marco se estudiaron las propiedades de figuras sumamente complicadas, incluyendo conjuntos de puntos. Por otro lado, apareció una actitud fundamentalmente nueva hacia las propiedades de las figuras en estudio. Se distinguieron distintos grupos de propiedades que podían estudiarse por separado, y esta misma abstracción *dentro* de la geometría dio lugar a muchas de sus ramas características, que se convirtieron en «geometrías» esencialmente independientes. El desarrollo de la geometría en todas estas direcciones continúa, y cada día se estudian nuevos «espacios» con sus correspondientes «geometrías»: el espacio de Lobachevski, el espacio proyectivo, el euclídeo y otros espacios de distintas dimensiones, en particular el espacio de cuatro dimensiones, los espacios de Riemann, los de Finsler, los espacios topológicos, etc. Estas teorías encuentran importantes aplicaciones no sólo en la geometría, sino en toda la matemática y también en la física y la mecánica; particularmente importantes son sus aplicaciones a la teoría de la relatividad de la física contemporánea, que es una teoría del espacio, del tiempo y de la gravitación. Es claro que el cambio que se ha operado en la geometría es un cambio cualitativo.

Las ideas de la geometría contemporánea y algunos de los elementos de la teoría de los distintos espacios que estudia se desarrollarán en los capítulos 16 y 17.

3. *Algebra*

El álgebra también experimentó un cambio cualitativo. En la primera mitad del siglo XIX surgieron nuevas teorías que condujeron a cambios en el carácter de aquélla, y a una extensión de su objeto y de su campo de aplicación.

En su forma original —como ya mencionamos en el §5— el álgebra se ocupaba de las operaciones matemáticas con números, considerados éstos desde un punto de vista formal, con abstracción de números concretos dados. Esta abstracción encontró su expresión en el hecho de que en el álgebra las magnitudes se notan por letras, sobre las que se realizan los cálculos de acuerdo con reglas formales bien establecidas.

El álgebra contemporánea retiene esta base, pero la amplía considerablemente. Ahora considera «magnitudes» de una naturaleza mucho más general que la de los números, y sobre estas «magnitudes» estudia operaciones que son bastante análogas en sus propiedades formales a las operaciones ordinarias de la aritmética: adición, substracción, multiplicación y división. Un ejemplo muy sencillo lo ofrecen las magnitudes vectoriales, que pueden ser sumadas por la conocida regla del paralelogramo. Pero la generalización realizada en el álgebra contemporánea es tal que incluso el término «magnitud» pierde a menudo su significado y se habla con mayor generalidad de «elementos» sobre los cuales es posible realizar operaciones similares a las operaciones algebraicas usuales. Por ejemplo, dos movimientos realizados uno a continuación del otro son evidentemente equivalentes a un cierto movimiento único, que es la suma de los dos; dos transformaciones algebraicas de una fórmula pueden ser equivalentes a una única transformación que produzca el mismo resultado, etc.; y así es posible hablar de la «adición» característica de los movimientos o de las transformaciones. Todo esto y mucho más se estudia de forma abstracta y general en el álgebra contemporánea.

Las nuevas teorías algebraicas en esta dirección surgieron en la primera mitad del siglo XIX gracias a las investigaciones de cierto número de matemáticos, entre los cuales debemos mencionar en particular al matemático francés Galois (1811-1832). Los conceptos, métodos y resultados del álgebra contemporánea encuentran importantes aplicaciones en el análisis, la geometría, la física y la cristalografía. En particular, la teoría que mencionamos al final del §3, concerniente a la simetría de los cristales y que fue desarrollada por E. S. Fedorov, se basa en la unión de la geometría con una de las nuevas teorías algebraicas, la llamada teoría de grupos.

Como se ve, nos encontramos ante una generalización cualitativa, fundamental, del objeto del álgebra, con un cambio del concepto mismo de la esencia del álgebra. Las ideas del álgebra contemporánea y los elementos básicos de algunas de sus teorías se explicarán en los capítulos 20 y 16.

4. *Análisis*

El análisis, en todas sus ramas, ha realizado también profundos progresos. En primer lugar, como ya se mencionó en la sección precedente, sus fundamentos adquirieron mayor rigor y surgieron nuevas definiciones generales y precisas de conceptos básicos tales como función, límite, integral y, finalmente, del concepto básico de magnitud variable (se dio una definición rigurosa de número real). El proceso de situar el análisis sobre una base más rigurosa fue iniciado por el matemático checo Bolzano (1781-1848), el matemático francés Cauchy (1789-1857) y otros. Este mayor rigor se logró al mismo tiempo que se hacían nuevos hallazgos en álgebra y geometría, y culminó en su forma actual en los años ochenta del siglo XIX gracias a los matemáticos alemanes Weierstrass, Dedekind y Cantor. Como se mencionó al final del §6, Cantor también puso los cimientos de la teoría de los conjuntos transfinitos, que desempeña un gran papel en el desarrollo de las novísimas ideas de la matemática.

La mayor precisión que adquirieron los conceptos de variable y función en conexión con la teoría de conjuntos, fue esencial para el posterior desarrollo del análisis. Se pasó al estudio de funciones más generales, y en esta misma línea se generalizó también el aparato del análisis, es decir, el cálculo diferencial e integral. Fue así como, a comienzos del siglo presente, surgió la nueva rama del análisis que ya mencionamos en el §6: la teoría de funciones de una variable real. El desarrollo de esta teoría va ligado principalmente a los matemáticos franceses Borel, Lebesgue y otros, así como a N. N. Luzin (1883-1950) y su escuela. En general, las ramas más recientes del análisis se incluyen bajo la rúbrica de análisis moderno, diferenciándolo así del antiguo, llamado análisis clásico.

En el análisis surgieron, además, otras teorías. Así, se formó una rama especial con la teoría de aproximación de funciones, que estudia los problemas relativos al mejor modo de representar aproximadamente funciones arbitrarias mediante funciones «simples», y en particular mediante polinomios, esto es, funciones de la forma

$$a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n.$$

La teoría de la aproximación de funciones tiene gran importancia, aunque sólo sea porque proporciona métodos generales para el cálculo práctico de funciones y para la sustitución aproximada de funciones complicadas por otras más sencillas. Los rudimentos de esta teoría aparecen ya en los comienzos mismos del análisis. La orientación actual se debe al gran matemático ruso P. L. Chebichev (1821-1894), orientación a partir de la cual se desarrolló más tarde la llamada teoría constructiva de funciones, gracias sobre todo a los trabajos de los matemáticos soviéticos, y en particular de S. N. Bernstein (nacido en 1880), a quien pertenecen los resultados más importantes en este campo. El capítulo 12 se ocupa de la aproximación de funciones.

Hablamos antes sobre el desarrollo de la teoría de funciones de una variable compleja. Debemos todavía mencionar la llamada teoría cualitativa de las ecuaciones diferenciales, que tuvo su origen en los trabajos de Poincaré (1854-1912) y de A. M. Lyapunov (1857-1918) y sobre la cual se darán algunas nociones en el capítulo 5, y también la teoría de las ecuaciones integrales. Estas teorías tienen gran importancia práctica en la mecánica, la física y la tecnología. Así, la teoría cualitativa de las ecuaciones diferenciales proporciona soluciones a problemas que conciernen a la estabilidad del movimiento, a la acción de mecanismos o de sistemas eléctricos vibrantes, y problemas similares. La estabilidad de un proceso significa en sentido general que, si se introducen pequeños cambios en los datos iniciales o en las condiciones del movimiento, entonces su transcurso sólo se modificará ligeramente. No hace falta hacer hincapié en la trascendencia técnica de cuestiones de este tipo.

5. *Análisis funcional*

Sobre la base proporcionada por el desarrollo del análisis y la física matemática, y junto con las nuevas ideas de la geometría y el álgebra, ha madurado una nueva y extensa sección de la matemática, el llamado análisis funcional, que tiene un papel excepcionalmente importante en la matemática moderna. Muchos matemáticos intervinieron en su creación; mencionemos, por ejemplo, al matemático alemán más importante de los tiempos actuales, Hilbert (1862-1943), al matemático húngaro Riesz (1880-1956) y al matemático polaco Banach (1892-1945). El capítulo 18 estará dedicado al análisis funcional.

La esencia de esta nueva rama de la matemática se puede resumir, en pocas palabras, del modo siguiente. Mientras que en el análisis

clásico la variable es una magnitud o «número», en el análisis funcional se considera como variable la función misma. Las propiedades de una función particular se determinan, no como tales propiedades, sino en relación con otras funciones. Lo que se estudia no es una función aislada sino toda una colección de funciones caracterizadas por una u otra propiedad; por ejemplo, la colección de todas las funciones continuas. Tal colección de funciones constituye lo que se denomina un espacio funcional. Este procedimiento corresponde, por ejemplo, al hecho de considerar la colección de todas las curvas sobre una superficie o de todos los posibles movimientos de un sistema mecánico dado, definiendo así las propiedades de las curvas o movimientos particulares en su relación con otras curvas o movimientos.

La transición de la investigación de funciones individuales a la investigación de una función *variable* es similar al paso de los números desconocidos x , y a las variables x , y , esto es, semejante a la idea de Descartes mencionada en un epígrafe anterior. En base a esta idea, Descartes desarrolló la célebre fusión del álgebra y la geometría: la fusión de una ecuación y una curva, que fue uno de los elementos más importantes en la aparición del análisis. De modo similar, la unión del concepto de función variable con las ideas del álgebra y la geometría contemporánea dieron lugar al nuevo análisis funcional. Igual que el análisis fue necesario para el desarrollo de la mecánica, así el análisis funcional suministró nuevos métodos para la solución de los problemas actuales de la física matemática y proporcionó el aparato matemático que precisaba la nueva mecánica cuántica del átomo. Como siempre, la historia se repite, pero en una nueva forma, en un plano superior. Como ya hemos dicho, el análisis funcional reúne las ideas básicas y los métodos del análisis, del álgebra moderna y de la geometría, y a su vez influye sobre el desarrollo de estas ramas de la matemática. Los problemas que se derivan del análisis clásico encuentran ahora, a menudo casi de golpe, nuevas soluciones más generales por medio del análisis funcional. Como en un foco se concentran aquí de un modo muy productivo las ideas más generales y abstractas de la matemática moderna.

De este breve esquema, de esta sencilla enumeración de las nuevas direcciones del análisis (la teoría de funciones de una variable real, la teoría de la aproximación de funciones, la teoría cualitativa de las ecuaciones diferenciales, la teoría de las ecuaciones integrales y el análisis funcional) puede verse que nos encontramos de hecho ante una etapa esencialmente nueva en el desarrollo del análisis.

6. *El cómputo matemático y la lógica matemática*

En todas las épocas el nivel técnico de los medios de cálculo ha ejercido una influencia esencial sobre los métodos matemáticos; pero los métodos de cálculo de que disponíamos hasta fecha muy reciente eran muy limitados. Los instrumentos más sencillos —tales como los ábacos, las tablas de logaritmos y la regla de cálculo—, la máquina de calcular y finalmente las calculadoras más complicadas y las máquinas de calcular automáticas, han sido las herramientas básicas de cálculo hasta los años cuarenta de este siglo. Estos instrumentos permitían realizar de un modo más o menos rápido las operaciones de adición, multiplicación, etc. Ahora bien, llegar al resultado numérico final de los problemas prácticos que surgen en nuestros días requiere un número colosal de tales operaciones, sucediéndose unas a otras en un complicado programa que depende a veces de los resultados obtenidos en el transcurso del cálculo. La resolución de tales problemas resultaba prácticamente imposible o del todo inútil debido a la larga duración del proceso de cálculo. Pero en los diez últimos años, se ha verificado un cambio radical en la ciencia del cálculo. Las modernas máquinas de calcular, construidas sobre la base de nuevos principios, permiten efectuar cálculos a una velocidad excepcionalmente grande y realizar al mismo tiempo complicadas cadenas de cálculos automáticamente, según programas muy flexibles preparados por adelantado.

Las nuevas técnicas no sólo permiten realizar investigaciones que hasta ahora eran de hecho casi impracticables, sino que también nos obligan a modificar nuestro juicio en punto al valor de muchos resultados matemáticos conocidos. Por ejemplo, dichas técnicas han dado un impulso especial al desarrollo de los métodos aproximados, esto es, métodos que permiten alcanzar, mediante una cadena de operaciones elementales, el resultado numérico deseado con una precisión suficientemente alta. Hoy día es preciso valorar los métodos matemáticos desde el punto de vista de su adaptabilidad a las máquinas correspondientes.

En íntima conexión con el desarrollo de las técnicas de cálculo está la lógica matemática. Esta se desarrolló, en principio, como resultado de dificultades intrínsecas surgidas en la misma matemática, y su objeto era el análisis de las demostraciones matemáticas. Es en sí misma una rama de la matemática, e incluye aquella rama de la lógica general que puede ser objetivamente formulada o desarrollada por métodos matemáticos.

Así, pues, aunque por un lado la lógica matemática se remonta a las mismas fuentes y fundamentos de la matemática, está, por otra parte, íntimamente relacionada con las cuestiones más actuales de las técnicas de cálculo. Naturalmente, una demostración, por ejemplo, que conduce a la puesta en marcha de un proceso concreto y que permite llegar al resultado deseado con un grado de precisión dado, es esencialmente diferente de las demostraciones abstractas en torno a la existencia de dicho resultado.

En relación con esto surge asimismo toda una serie de cuestiones concernientes al grado de generalidad que se puede alcanzar en problemas que pueden resolverse por métodos que están completamente definidos de antemano en cada una de sus etapas. En este terreno se han alcanzado resultados profundos en lógica matemática, resultados que son extremadamente importantes desde el punto de vista epistemológico general.

No sería exagerado decir que con el desarrollo de las nuevas técnicas de cálculo y los resultados de la lógica matemática ha comenzado un nuevo período en la matemática moderna, caracterizado por el hecho de que su objetivo no es sólo el estudio de determinados entes, sino también de los caminos y medios por los cuales dichos entes pueden ser definidos; no sólo determinados problemas, sino también todos los posibles métodos de resolverlos.

A lo antedicho sólo es preciso añadir que también las ramas más antiguas de la matemática —la teoría de números, la geometría euclídea, el álgebra y el análisis clásico, y la teoría de probabilidades— han experimentado un rápido desarrollo en todo el período de la matemática moderna, por lo cual estos campos se han enriquecido con muchas ideas y resultados nuevos y fundamentales; mencionemos, por ejemplo, los resultados obtenidos en teoría de números y en la geometría del espacio usual por los matemáticos rusos y soviéticos P. L. Chebichev, E. S. Fedorov, I. M. Vinogradov, y otros. El desarrollo, a lo largo de un amplio frente, de la teoría de las probabilidades ha estado relacionado con las importantísimas regularidades que se observan en la física estadística y en los problemas contemporáneos de la tecnología.

7. Rasgos característicos de la matemática moderna

¿Cuáles son las características más generales del conjunto de la matemática moderna que la distinguen de los anteriores desarrollos de la geometría, el álgebra y el análisis?

En primer lugar, la inmensa extensión de la materia que cubren

la matemática y sus aplicaciones. Tal extensión del objeto y del campo de aplicación representa un enorme crecimiento cuantitativo y cualitativo, originado por la aparición de nuevas y poderosas teorías y métodos que permiten resolver problemas completamente inaccesibles hasta ahora. Esta extensión del objeto de la matemática está caracterizada por el hecho de que la matemática contemporánea se ha impuesto conscientemente a sí misma la tarea de estudiar todos los tipos posibles de relaciones cuantitativas y formas espaciales.

Un segundo rasgo característico de la matemática moderna es la formación de conceptos generales a un nuevo y más alto nivel de abstracción. Es precisamente este rasgo el que garantiza la preservación de la unidad de la matemática, a pesar de su inmenso crecimiento en ramas muy diferentes. Incluso en partes de la matemática que se encuentran en extremo desligadas unas de otras se observan similitudes de estructura cuyo origen descansa en los conceptos y teorías generales de nuestros días. Ello garantiza a los métodos matemáticos contemporáneos una gran generalidad y amplitud en sus aplicaciones, y en particular lleva a una profunda interpenetración de las ramas fundamentales de la matemática: la geometría, el álgebra y el análisis.

Como otro de los rasgos característicos de la matemática moderna debemos mencionar también la evidente preponderancia del punto de vista de la teoría de conjuntos. Naturalmente, este punto de vista debe su importancia al hecho de que resume en cierto sentido el rico contenido de todos los desarrollos precedentes de la matemática. Finalmente, uno de los rasgos más característicos de la matemática moderna es el análisis profundo de sus fundamentos, de la influencia mutua de sus conceptos, de la estructura de sus diversas teorías y de los métodos de demostración matemática. Sin un tal análisis de los fundamentos no sería posible mejorar ni seguir desarrollando los principios y las teorías que han conducido a las generalizaciones actuales.

El rasgo característico de la matemática moderna puede expresarse de la siguiente manera: su objeto no sólo consta de relaciones y formas cuantitativas dadas, sino de todas las posibles. En geometría no se habla únicamente de las relaciones y formas espaciales, sino de todas las formas posibles semejantes a las espaciales. En álgebra hablamos de distintos sistemas abstractos de objetos con todas las posibles operaciones que se pueden efectuar sobre ellos. En análisis no sólo se consideran como variables las magnitudes, sino también las propias funciones. En un espacio funcional se encuentran todas

las funciones de un tipo dado (todas las posibles interdependencias entre las variables). Resumiendo, es posible decir que, mientras la matemática elemental se ocupa de las magnitudes constantes, y el siguiente período de las magnitudes variables, *la matemática contemporánea es la matemática de todas las posibles relaciones e interdependencias cuantitativas (en general variables) entre magnitudes*. Esta definición es, naturalmente, incompleta, pero pone el acento en el rasgo característico de la matemática moderna, rasgo que la distingue de la de otras épocas ²⁵.

²⁵ Esta sección va seguida en el texto original ruso de dos secciones tituladas «La naturaleza esencial de la matemática» y «Las leyes del desarrollo de la matemática». Estas secciones se omiten en la presente traducción en vista del hecho de que discuten con mayor detalle, y desde el punto de vista filosófico general del materialismo dialéctico, puntos de vista ya establecidos con gran claridad en las secciones precedentes. (Nota del editor americano.)

Capítulo 2

ANALISIS

§1. *Introducción*

La aparición, a finales de la Edad Media, de una nueva situación industrial en Europa —como fue el nacimiento del capitalismo, que vino a reemplazar al sistema feudal— fue acompañada de importantes descubrimientos geográficos y exploraciones. En 1492, basándose en la idea de que la Tierra era esférica, Colón descubrió el Nuevo Mundo. Este descubrimiento amplió enormemente las fronteras del mundo conocido y determinó una revolución en el pensamiento de la humanidad. El final del siglo xv y el comienzo del xvi vieron la actividad creadora de los grandes artistas-humanistas Leonardo da Vinci, Rafael y Miguel Angel, que dieron un nuevo significado al arte. En 1543 Copérnico publicó su trabajo «Sobre el movimiento de los cuerpos celestes», que modificó por completo la ciencia de la astronomía; en 1609 apareció la «Nueva astronomía» de Kepler, que contenía su primera y segunda ley sobre el movimiento de los planetas alrededor del Sol, y en 1618 su libro «La armonía del mundo», que contenía la tercera ley. Basándose en el estudio de los trabajos de Arquímedes y en sus propios y audaces ex-

perimentos, Galileo puso los cimientos de la mecánica, ciencia indispensable para la nueva tecnología. En 1609 Galileo dirigió su nuevo telescopio —todavía pequeño y rudimentario— hacia los cielos, y la primera visión fue suficiente para destruir las esferas celestes ideales de Aristóteles y el dogma de la forma perfecta de los cuerpos celestes. Se comprobó que la superficie de la Luna estaba cubierta de montañas y cráteres. Venus manifestaba fases como la Luna; Júpiter estaba rodeado por cuatro satélites y suministraba un modelo en miniatura del sistema solar. La Vía Láctea quedó desglosada en estrellas y, por primera vez, los hombres intuyeron las inmensas distancias estelares. Ningún otro descubrimiento científico había producido tal impresión sobre el mundo civilizado¹.

El desarrollo que experimentó la navegación —y como consecuencia de ello también la astronomía— y el nuevo desarrollo de la tecnología y la mecánica precisaban del estudio de muchos problemas matemáticos nuevos. La novedad de estos problemas consistía principalmente en el hecho de que requerían el estudio matemático de las leyes del movimiento, en el amplio sentido de esta expresión.

El estado de reposo e inmovilidad es desconocido en la naturaleza. Todas las cosas de la naturaleza, desde las más pequeñas partículas hasta los cuerpos de mayor masa, están en un estado de eterna creación y aniquilación, en un flujo perpetuo, en un movimiento y cambio incesantes. En último término, toda ciencia natural estudia algún aspecto de este movimiento. El análisis matemático es la rama de la matemática que proporciona métodos para la investigación cuantitativa de los distintos procesos de cambio, movimiento y dependencia de una magnitud respecto de otras. Surge así, de manera natural, en un período en el que el desarrollo de la mecánica y la astronomía nacidas de los problemas de la tecnología y la navegación, habían proporcionado ya un cúmulo considerable de observaciones, medidas e hipótesis y estaban impulsando a la ciencia hacia la investigación cuantitativa de las formas más sencillas del movimiento.

El nombre «análisis infinitesimal» no dice nada sobre el objeto de estudio, sino que enfatiza el método. Se trata del método matemático especial de los infinitésimos o, en su forma moderna, de los límites. Damos ahora algunos ejemplos típicos de razonamientos que hacen uso del método de los límites y en una de las últimas secciones definiremos los conceptos necesarios.

¹ Esta sección está basada en el bello ensayo del académico S. I. Vavilov «Galileo» (Gran enciclopedia soviética, volumen 10, 1952).

Ejemplo 1

Como estableció experimentalmente Galileo, la distancia s recorrida en el tiempo t por un cuerpo que cae libremente en el vacío viene expresada por la fórmula

$$s = \frac{gt^2}{2} \quad (1)$$

(g es una constante igual a $9,81 \text{ m/seg}^2$)². ¿Cuál es la velocidad de un cuerpo que cae libremente en cada punto de su trayectoria?

Supongamos que el cuerpo pasa por el punto A en el tiempo t y consideremos lo que sucede en el breve intervalo de tiempo Δt ; esto es, en el intervalo que va desde t a $t + \Delta t$. La distancia recorrida aumentará en un cierto valor Δs . La distancia original es $s_1 = gt^2/2$; la distancia incrementada es

$$s_2 = \frac{g(t + \Delta t)^2}{2} = \frac{gt^2}{2} + \frac{g}{2} (2t \Delta t + \Delta t^2).$$

De aquí se deduce que:

$$\Delta s = s_2 - s_1 = \frac{g}{2} (2t \Delta t + \Delta t^2).$$

Δs representa la distancia recorrida en el intervalo de tiempo Δt . Para encontrar la velocidad media en el intervalo Δs , dividimos Δs por Δt :

$$v_m = \frac{\Delta s}{\Delta t} = gt + \frac{g}{2} \Delta t.$$

Haciendo que Δt tienda a cero obtenemos una velocidad media que se aproxima tanto como queramos a la velocidad verdadera en el punto A . Por otra parte, vemos que el segundo sumando del miembro de la derecha de la ecuación se hace arbitrariamente pequeño

² Actualmente la fórmula (1) se deduce de las leyes generales de la mecánica, pero históricamente fue justamente esta fórmula la que, después de haber sido establecida experimentalmente por Galileo, sirvió como una de las experiencias que fueron más tarde generalizadas por estas leyes.

cuando decrece Δt , por lo que el valor de v_m se aproxima al valor gt , hecho que se puede escribir como sigue:

$$v = \lim_{\Delta t \rightarrow 0} v_m = \lim_{\Delta t \rightarrow 0} \frac{\Delta s}{\Delta t} = \lim_{\Delta t \rightarrow 0} \left(gt + \frac{g}{2} \Delta t \right) = gt.$$

Según esto, gt es la velocidad verdadera en el instante t .

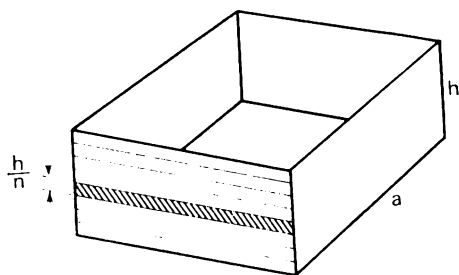


FIG. 1

Ejemplo 2

Un recipiente de base cuadrada de lado a y paredes verticales de altura h está lleno de agua hasta el borde (Fig. 1). ¿Qué fuerza ejerce el agua sobre cada una de las paredes del recipiente?

Dividimos la superficie de la pared en n rectángulos horizontales de altura h/n . La presión ejercida en cada punto del recipiente es igual, según una conoci-

da ley, al peso de la columna de agua que se encuentra encima de él. Así, en el borde inferior de cada uno de los rectángulos la presión, expresada en unidades convenientes, será respectivamente igual a

$$\frac{h}{n}, \frac{2h}{n}, \frac{3h}{n}, \dots, \frac{(n-1)h}{n}, h.$$

Para obtener una expresión aproximada de la fuerza P supondremos que la presión es constante dentro de cada rectángulo. Así el valor aproximado de P es igual a

$$\begin{aligned} P &\approx \frac{ah}{n} \cdot \frac{h}{n} + \frac{ah}{n} \cdot \frac{2h}{n} + \dots + \frac{ah}{n} \cdot \frac{(n-1)h}{n} + \frac{ah}{n} h = \\ &= \frac{ah^2}{n^2} (1 + 2 + \dots + n) = \frac{ah^2}{n^2} \cdot \frac{n(n+1)}{2} = \frac{ah^2}{2} \left(1 + \frac{1}{n} \right). \end{aligned}$$

Para encontrar el verdadero valor de la fuerza dividimos la pared en rectángulos cada vez más estrechos, haciendo crecer n indefinidamente. Al aumentar n , el valor $1/n$ de la fórmula anterior se hará más y más pequeño, y en el límite obtenemos la fórmula exacta

$$P = \frac{ah^2}{2}.$$

La idea del método del límite es sencilla y equivale a lo siguiente. Para determinar el valor exacto de una cierta magnitud determinamos primero, no la magnitud en sí, sino una aproximación a ella. Sin embargo, no hacemos una única aproximación sino una serie de ellas, cada una de las cuales es más precisa que la anterior. Del examen de esta cadena de aproximaciones, esto es, del examen del proceso mismo de aproximación, determinamos unívocamente el valor exacto de la magnitud. Por este método, que es en esencia profundamente dialéctico, obtenemos una constante fija como resultado de un proceso o movimiento.

El método matemático del límite fue el fruto de la persistente labor de muchas generaciones sobre problemas que no podían resolverse por los sencillos métodos de la aritmética, el álgebra y la geometría elemental.

¿Cuáles eran los problemas cuya solución conducía a los conceptos fundamentales del análisis, y cuáles eran los métodos de resolución que se utilizaban en estos problemas? Examinemos algunos de ellos.

Los matemáticos del siglo xvii se fueron percatando gradualmente de que una gran parte de los problemas que surgían de distintos tipos de movimiento (con la consiguiente dependencia de unas variables respecto a otras), así como de problemas geométricos que no se habían podido abordar con los métodos usuales, podían reducirse a dos tipos. Ejemplos sencillos de problemas del primer tipo son: hallar la velocidad en cualquier instante de un movimiento no uniforme (o, en general, encontrar la velocidad de variación de una magnitud dada), y trazar una tangente a una curva dada. Estos problemas (el primer ejemplo es uno de ellos) condujeron a una rama del análisis que recibió el nombre de «cálculo diferencial». Ejemplos muy sencillos del segundo tipo de problemas son: encontrar el área de una figura curvilínea (el problema de la cuadratura), o la distancia recorrida en un movimiento no uniforme, o, en general, el efecto total de la acción de una magnitud continuamente variable (compá-

rese con el segundo ejemplo de arriba). Este grupo de problemas condujo a otra rama del análisis, el «cálculo integral». Así se concretizaron dos problemas fundamentales: el problema de las tangentes y el problema de las cuadraturas.

En este capítulo describiremos con detalle las ideas que sirvieron de base para la solución de estos dos problemas. Particularmente importante es el teorema de Newton y Leibnitz; el problema de las cuadraturas es, en cierto sentido, el inverso del problema de las tangentes. Para resolver el problema de las tangentes y aquellos otros que pueden reducirse a él se inventó un algoritmo conveniente, un método completamente general que conduce directamente a la solución: el método de las derivadas o de la diferenciación.

La historia de la creación y el desarrollo del análisis y del papel que tuvo en su crecimiento la geometría analítica de Descartes ya han sido descritos en el capítulo 1. Vimos que en la segunda mitad del siglo XVII y en la primera del XVIII tuvo lugar un cambio radical en el conjunto de la matemática. A las disciplinas que ya existían —aritmética, geometría elemental y algunos rudimientos de álgebra y trigonometría— se añadieron métodos generales tales como la geometría analítica, el cálculo diferencial e integral y la teoría de las ecuaciones diferenciales más sencillas, lo cual permitió resolver problemas cuyas soluciones habían sido hasta entonces inaccesibles.

Se encontró que, si la ley de formación de una curva no es demasiado complicada, entonces siempre es posible trazar la tangente en cualquier punto dado; basta para ello calcular, con la ayuda del cálculo diferencial, la derivada, lo cual en la mayoría de los casos requiere muy poco tiempo. Hasta entonces sólo se sabía trazar tangentes al círculo y a una o dos curvas más, sin sospechar la existencia de una solución general para este problema.

Si conocemos la distancia recorrida por un punto en movimiento en cualquier instante de tiempo, utilizando ese mismo método podemos encontrar en seguida la velocidad del punto en un instante dado, y también su aceleración. Gracias a esta técnica no fue muy difícil, por ejemplo, demostrar a partir de las leyes newtonianas del movimiento y de la ley de la gravitación universal que los planetas deben moverse alrededor del Sol según elipses de acuerdo con las leyes de Kepler.

De gran importancia práctica es el problema de obtener los valores mayor y menor de una magnitud: el problema de los máximos y mínimos. Tomemos un ejemplo: de un tronco de árbol de sección transversal circular y radio dado deseamos cortar un tablón de sec-

ción transversal rectangular tal que ofrezca la mayor resistencia posible a curvarse. ¿En qué proporción deberán estar sus lados? Un breve razonamiento sobre la rigidez de los tabloncillos de sección transversal rectangular (aplicando conceptos elementales del cálculo integral), seguido por la resolución de un problema de máximos (que implica el cálculo de una derivada), da como solución que la mayor rigidez se obtiene para una sección transversal rectangular cuya altura está en la proporción $\sqrt{2} : 1$ respecto a la base. Los problemas de máximos y mínimos son de resolución tan sencilla como los de trazar tangentes.

La curvatura en distintos puntos de una línea curva (salvo en la recta y el círculo) es en general diferente. ¿Cómo calcular el radio de un círculo que tenga la misma curvatura que la curva problema en el punto dado, esto es, cuál es el radio de curvatura de la curva en dicho punto? Este problema es también muy sencillo; basta aplicar dos veces la operación de derivación. El radio de curvatura es de gran importancia en muchas cuestiones de mecánica.

Antes de la invención de los nuevos métodos de cálculo sólo era posible calcular el área de los polígonos, del círculo, de un sector o de un segmento circular, y de dos o tres figuras más. Por otro lado, Arquímedes ya había inventado un método para calcular el área de un segmento de parábola. El ingenioso método que empleó en este problema se basaba en propiedades especiales de la parábola y llevó a la sospecha de que todo nuevo problema de cálculo de áreas debería igualmente requerir un método propio de investigación, incluso más ingenioso y difícil que el de Arquímedes. Así, los matemáticos se vieron agradablemente sorprendidos cuando se comprobó que el problema de Newton y Leibnitz, según el cual la inversión del problema de las tangentes resolvía el problema de la cuadratura, suministraba un método directo para calcular las áreas limitadas por curvas muy variadas. Se trataba de un método general, aplicable a un número infinito de figuras muy diversas. La misma observación es válida para el cálculo de volúmenes, superficies, longitudes de curvas, masas de cuerpos no homogéneos, etc.

El nuevo método se mostró más fructífero aún en el terreno de la mecánica. Parecía que no había problema en la mecánica que el nuevo cálculo no pudiera aclarar y resolver.

Poco antes Pascal había explicado el aumento del vacío de Torricelli con la altitud como una consecuencia del decrecimiento de la presión atmosférica. Pero, ¿cuál es la ley que gobierna este decre-

cimiento? La cuestión se resuelve inmediatamente estudiando una sencilla ecuación diferencial.

Es bien sabido de los navegantes que para que un solo hombre sea capaz de mantener amarrado un gran navío, es preciso dar dos o tres vueltas al cable de amarrar alrededor del cabrestante. ¿Por qué es esto? Se vio que desde un punto de vista matemático el problema es casi idéntico al precedente y puede resolverse al instante.

Así, después de la aparición del análisis siguió un período en el que éste fue aplicado a las más variadas cuestiones de la tecnología y las ciencias naturales. Puesto que se funda en la abstracción de los rasgos específicos de problemas particulares, el análisis matemático refleja las propiedades reales que subyacen en el mundo material; y esta es la razón por la cual proporciona los medios para la investigación de cuestiones prácticas de tan diversas características. El movimiento de los cuerpos sólidos, el de las partículas en los líquidos y gases, sus leyes de flujo, la conducción del calor y la electricidad, las reacciones químicas, todos estos fenómenos se estudian en las correspondientes ciencias con los medios del análisis matemático.

Al mismo tiempo que se extendían sus aplicaciones, el propio análisis se enriqueció notablemente con la creación y desarrollo de nuevas ramas, tales como la teoría de las series, las aplicaciones de la geometría al análisis y la teoría de las ecuaciones diferenciales.

Entre los matemáticos del siglo XVIII prevalecía la opinión de que cualquier problema de las ciencias naturales podía resolverse por medio de la geometría analítica y del cálculo diferencial, con tal solamente que se pudiera encontrar una descripción matemática correcta de él.

Los matemáticos llegaron gradualmente a problemas más complicados de las ciencias naturales y la tecnología, problemas que necesitaban de un mayor desarrollo de sus métodos y para cuya solución fue preciso crear otras ramas de la matemática: el cálculo de variaciones, la teoría de funciones de una variable compleja, la teoría de campos, las ecuaciones integrales y el análisis funcional. Pero todos estos nuevos métodos de cálculo eran en esencia extensiones y generalizaciones inmediatas de los notables métodos descubiertos en el siglo XVII. Los más grandes matemáticos del siglo XVIII —David Bernoulli (1700-1782), Leonard Euler (1707-1783) y Lagrange (1736-1813)—, que abrieron nuevos caminos a la ciencia, tomaron constantemente como punto de partida problemas fundamentales de las ciencias físicas. Este gran desarrollo del análisis continuó en el siglo XIX gracias a prestigiosos matemáticos como Gauss (1777-1855),

Cauchy (1789-1857), M. V. Ostrogradski (1801-1861), P. L. Chebichev (1821-1894), Riemann (1826-1866), Abel (1802-1829), Weierstrass (1815-1897), todos los cuales hicieron contribuciones verdaderamente notables al desarrollo del análisis matemático.

El genial matemático ruso N. I. Lobachevski influyó en el desarrollo de ciertas cuestiones del análisis matemático, siendo también dignos de mencionar los principales matemáticos de principios de este siglo: A. A. Markov (1856-1922), A. M. Lyapunov (1857-1918), H. Poincaré (1854-1912), F. Klein (1849-1925), D. Hilbert (1862-1943).

La segunda mitad del siglo XIX presenció un profundo examen crítico y el esclarecimiento de los fundamentos del análisis. Los distintos y poderosos métodos que se habían ido acumulando fueron sistematizados de acuerdo con el creciente nivel de rigor matemático. Todos estos métodos constituyen la herramienta con que (junto con la aritmética, el álgebra, la geometría y la trigonometría) damos una interpretación matemática al mundo que nos rodea, describimos el curso de los sucesos reales y resolvemos importantes problemas prácticos relacionados con ellos.

La geometría analítica, el cálculo diferencial e integral y la teoría de las ecuaciones diferenciales se estudian en todas las escuelas técnicas superiores, lo que hace que estas ramas de la matemática sean conocidas por millones de ciudadanos; estas ciencias se enseñan también, a nivel elemental, en muchas escuelas técnicas de grado medio, y hay intención de introducirlas asimismo en el bachillerato.

En época muy reciente, el uso generalizado de máquinas rápidas de calcular ha introducido una nueva era en la matemática. Estas máquinas, junto con las ramas de la matemática que acabamos de mencionar, abren nuevas y fantásticas posibilidades al género humano.

En el momento actual el análisis, y todas las ramas que surgen de él, representan una ciencia matemática ampliamente diversificada, consistente en varias disciplinas íntimamente conectadas las unas con las otras; el desarrollo y perfeccionamiento de cada una de estas disciplinas continúa en la actualidad.

El papel que desempeñan hoy día las necesidades prácticas y los problemas relacionados con el desarrollo tecnológico en el análisis es más importante que nunca. De gran importancia son los problemas aerodinámicos de las velocidades supersónicas, que están siendo resueltos con enorme éxito. Los problemas más arduos de la física matemática han alcanzado ahora la etapa en la que pueden ser resueltos en forma numérica. En la física contemporánea, teorías tales como la mecánica cuántica (que estudia los problemas relativos

al microcosmos del átomo) no sólo requieren el concurso de las ramas más avanzadas del análisis matemático contemporáneo para resolver sus problemas, sino que incluso sus conceptos fundamentales no podrían ser descritos sin el análisis.

El objeto del presente capítulo es dar una presentación adecuada, asequible para el lector que sólo conozca la matemática elemental, del desarrollo y de las aplicaciones más sencillas de los conceptos básicos del análisis, tales como función, límite, derivada e integral. Puesto que las distintas ramas específicas del análisis se tratarán en otras secciones del libro, el presente capítulo tiene un carácter más elemental; el lector que ya haya estudiado un primer curso de análisis puede, por tanto, omitirlo sin peligro de no entender el resto del libro.

§2. *Función*

El concepto de función

Los distintos objetos y fenómenos que observamos en la naturaleza están orgánicamente relacionados unos con otros; son interdependientes. El género humano conoce desde hace tiempo las relaciones más sencillas de esta clase, y este conocimiento se halla expresado en las leyes físicas. Estas leyes indican que las distintas magnitudes que caracterizan un fenómeno dado están tan íntimamente relacionadas que algunas de ellas quedan completamente determinadas por los valores de las demás. Por ejemplo, la longitud de los lados de un rectángulo determina completamente su área; el volumen de una cantidad conocida de gas a temperatura dada viene determinado por la presión, y el alargamiento de una varilla metálica está determinado por su temperatura. Fueron correspondencias de esta clase las que sirvieron de origen al concepto de *función*.

Consideremos una fórmula algebraica que, a cada valor de las magnitudes literales que aparecen en ella, haga corresponder un valor de la magnitud expresada por la fórmula; ésta es la idea básica de función. Veamos algunos ejemplos de funciones expresadas por tales fórmulas.

1. Supongamos que en un cierto instante un punto material que estaba en reposo comienza a caer por la acción de la gravedad. La distancia recorrida por el punto en el tiempo t está expresada por la fórmula

$$s = \frac{gt^2}{2}, \quad (1)$$

donde g es la aceleración de la gravedad.

2. A partir de un cuadrado de lado a construimos una caja rectangular abierta de altura x (Fig. 2). El volumen V de la caja se calcula por la fórmula

$$V = x(a - 2x)^2. \quad (2)$$

Para cada altura x , y con la restricción evidente $0 \leq x \leq a/2$, la fórmula (2) permite encontrar el volumen de la caja.

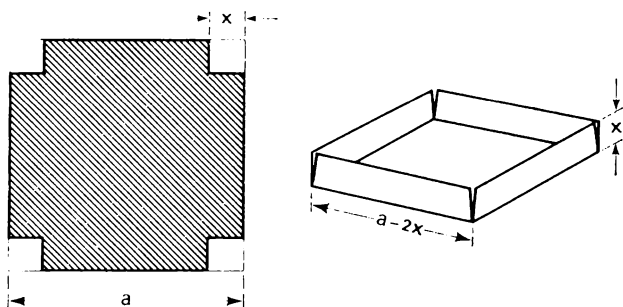


FIG. 2

3. Consideremos una farola (Fig. 3) situada en el centro de una pista de patinaje circular con una luz a la altura h . La iluminación T en el borde del círculo puede expresarse por la fórmula

$$T = \frac{A \operatorname{sen} \alpha}{h^2 + r^2}, \quad (3)$$

donde r es el radio del círculo, $\operatorname{tg} \alpha = h/r$, y A es una cierta magnitud que caracteriza la intensidad de la luz. Si conocemos la altura h podemos calcular T por la fórmula (3).

4. Las raíces de la ecuación de segundo grado

$$x^2 + px - 1 = 0 \quad (4)$$

vienen dadas por la fórmula

$$x = -\frac{p}{2} \pm \sqrt{1 + \frac{p^2}{4}}. \quad (5)$$

El rasgo característico de una fórmula en general, y de los ejemplos dados en particular, es que, para cualquier valor dado de una de las variables (denominada variable independiente: el tiempo t , la altura x de la caja, la altura h de la farola, el coeficiente p de la ecuación de segundo grado), permite calcular el valor de la otra variable (la distancia s , el volumen V , la iluminación T , la raíz x de la ecuación), que se llama variable dependiente o función de la primera variable.

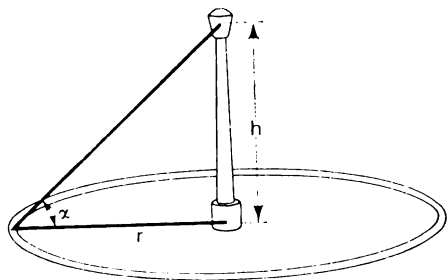


FIG. 3

Cada una de las fórmulas anteriores proporciona un ejemplo de función: la distancia s recorrida por el punto es función del tiempo t ; el volumen V de la caja es función de la altura x ; la iluminación T del borde de la pista es función

de la altura h de la farola; las dos raíces de la ecuación de segundo grado (4) son funciones del coeficiente p .

Debe observarse que en algunos casos la variable independiente puede tomar cualquier valor numérico posible, como en el ejemplo 4, donde el coeficiente p de la ecuación de segundo grado (4) puede ser cualquier número. En otros casos, la variable independiente sólo puede tomar valores de un conjunto (o colección) de números determinados de antemano, como en el ejemplo 2, donde el volumen de la caja es función de su altura x y ésta puede tomar cualquier valor del conjunto de números que satisfacen la desigualdad $0 < x < a/2$. Igualmente, en el ejemplo 3, la iluminación T en el borde de la pista es función de la altura h de la farola, la cual teóricamente puede tomar cualquier valor que satisfaga la desigualdad $h > 0$; pero en la práctica h debe satisfacer las desigualdades $0 < h \leq H$, donde la magnitud H viene determinada por los medios técnicos de que disponga el empresario de la pista.

Introduzcamos otros ejemplos de esta clase. La fórmula

$$y = \sqrt{1 - x^2}$$

define una función real (es decir, una relación entre los números reales x e y) sólo para aquellos valores de x que satisfacen las desigualdades $-1 \leq x \leq +1$; y la fórmula $y = \log(1 - x^2)$, únicamente para aquellos x que satisfacen las desigualdades $-1 < x < +1$.

Así, pues, es necesario tener en cuenta el hecho de que las funciones pueden no estar definidas para todos los valores numéricos de la variable independiente, sino sólo para aquellos que pertenecen a un cierto conjunto, que casi siempre llena un intervalo del eje x , con o sin sus puntos extremos.

Estamos ahora en condiciones de dar la definición de función aceptada por la matemática actual.

La variable (dependiente) y es función de la variable (independiente) x si existe una regla por la cual a cada valor de x, perteneciente a un cierto conjunto de números, corresponde un valor definido de y.

El conjunto de los valores x que aparece en esta definición se llama *dominio de la función*.

Todo nuevo concepto da lugar a un nuevo simbolismo. La transición de la aritmética al álgebra fue posible gracias a la construcción de fórmulas que eran válidas para números arbitrarios, y la búsqueda de soluciones generales dio lugar al simbolismo literal del álgebra.

El problema del análisis es el estudio de las funciones, esto es, de la dependencia de una variable respecto de otra. Por eso, y al igual que en álgebra tuvo lugar la transición de los números concretos a los números en general, denotados por letras, en análisis hubo la correspondiente transición de las fórmulas concretas a fórmulas generales. La frase « y es función de x » se escribe convencionalmente

$$y = f(x).$$

Lo mismo que en álgebra se emplean distintas letras para diferentes números, también en análisis se utilizan diferentes notaciones para los distintos tipos de dependencia, esto es, para funciones diferentes: así escribimos $y = F(x)$, $y = \phi(x)$, ...

Gráfica de una función

Una de las ideas más fructíferas y brillantes de la segunda mitad del siglo XVII fue la de la conexión entre el concepto de función y la representación geométrica de una curva. Esta conexión puede realizarse, por ejemplo, por medio de un sistema de coordenadas cartesianas rectangulares, con el cual el lector está ciertamente familiarizado de un modo general desde la matemática de bachillerato.

Situemos sobre el plano un sistema de coordenadas cartesianas rectangulares. Esto significa que sobre el plano elegimos dos rectas mutuamente perpendiculares (el eje de abscisas y el eje de ordenadas) y que sobre cada una de ellas fijamos una dirección positiva. A cada punto M del plano le podemos asignar entonces dos números (x, y) , que son sus coordenadas y que expresan, en el sistema de medida adoptado, la distancia (tomada con su signo³) del punto M a los ejes de ordenadas y abscisas, respectivamente.

Con un sistema tal de coordenadas podemos representar gráficamente las funciones en forma de ciertas líneas. Supongamos que nos dan una función

$$y = f(x). \quad (6)$$

Esto significa, como sabemos, que para todo valor de x perteneciente al dominio de definición de la función dada es posible determinar de algún modo, por medio del cálculo, por ejemplo el valor correspondiente de y . Damos ahora a x todos los valores numéricos posibles, determinamos y para cada x de acuerdo con la regla (6) y construimos sobre el plano el punto de coordenadas x e y . De este modo, a todo punto M' sobre el eje x (Fig. 4) le corresponderá un punto M

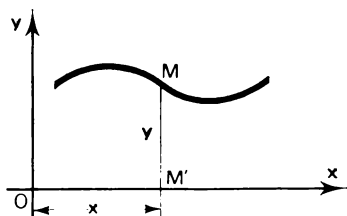


FIG. 4

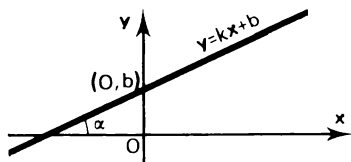


FIG. 5

de coordenadas x e $y = f(x)$. El conjunto de todos los puntos M forma una cierta línea, que se llama gráfica de la función $y = f(x)$.

Así, la gráfica de la función $f(x)$ es el lugar geométrico de los puntos cuyas coordenadas satisfacen la ecuación (6).

En los centros de segunda enseñanza se aprende normalmente a manejar las gráficas de las funciones más sencillas. Así el lector probablemente sabe que la función $y = kx + b$, donde k y b son constantes, tiene por gráfica (Fig. 5) una línea recta que forma un

³ El número x es la abscisa, e y , la ordenada del punto M .

ángulo α con la dirección positiva del eje x (siendo $\operatorname{tg} \alpha = k$) y que corta al eje y en el punto $(0, b)$. Esta función se llama *función lineal*.

Las funciones lineales aparecen con mucha frecuencia en las aplicaciones prácticas. Recordemos que muchas leyes físicas vienen representadas, con notable precisión, por funciones lineales. Por ejemplo, la longitud l de un cuerpo puede considerarse con buena aproximación como función lineal de su temperatura:

$$l = l_0 + \alpha l_0 t,$$

donde α es el coeficiente de dilatación lineal, y l_0 la longitud del cuerpo para $t = 0$. Si x es el tiempo e y la distancia recorrida por un punto en movimiento, entonces la función lineal $y = kx + b$ expresa evidentemente que el punto se mueve con velocidad uniforme k ; y el número b denota la distancia, en el instante $x_0 = 0$, del punto en movimiento, respecto del punto cero a partir del cual medimos las distancias. Las funciones lineales son extremadamente útiles por su sencillez y porque es posible considerar ciertas variaciones no uniformes como aproximadamente lineales, aunque sólo sea en pequeños intervalos.

Pero en muchos casos es necesario hacer uso de una dependencia funcional no lineal. Recordemos por ejemplo la ley de Boyle-Mariotte

$$v = \frac{c}{p},$$

donde las magnitudes p y v son inversamente proporcionales. La gráfica de tal relación es una hipérbola (Fig. 6).

La ley física de Boyle-Mariotte corresponde en realidad al caso en el que p y v son positivos; es decir, representa la rama de la hipérbola situada en el primer cuadrante.

La clase general compuesta por los procesos oscilatorios incluye movimientos periódicos que vienen normalmente descritos por las funciones trigonométricas. Por ejemplo, si colgamos un muelle del techo y lo estiramos a partir de su posición de equilibrio sin rebasar los límites elásticos del muelle, entonces el punto A realizará oscilaciones verticales que vendrán expresadas con bastante precisión por la ley

$$x = a \cos(pt + \alpha),$$

donde x es el desplazamiento del punto A a partir de su posición de equilibrio, t es el tiempo, y los números a , p y α son constantes que dependen del material, las dimensiones y la longitud inicial del muelle.

Hay que tener en cuenta que una función puede venir definida por diversas fórmulas según sea el dominio elegido. Por ejemplo, la relación $Q = f(t)$ entre la temperatura t de un gramo de agua (o hielo) y la cantidad de calor Q contenida en él, cuando t varía entre -10° y $+10^\circ$, es una función perfectamente determinada que

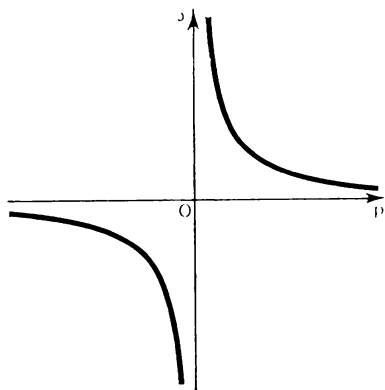


FIG. 6

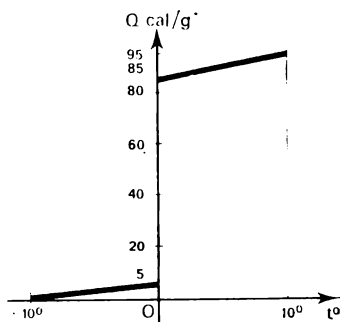


FIG. 7

resulta difícil de expresar en una única fórmula ⁴, pero fácil de representar mediante dos. Puesto que el calor específico del hielo es igual a 0,5 y el del agua es igual a 1, esta función (tomando $Q = 0$ a -10°) se representa por la fórmula:

$$Q = 0,5t + 5,$$

cuando t varía en el intervalo $-10^\circ \leq t < 0^\circ$, y por la fórmula

$$Q = t + 85,$$

cuando t varía en el intervalo $0^\circ < t \leq 10^\circ$. Para $t = 0$ esta función es indefinida o de valores múltiples; por conveniencia, podemos suponer que en $t = 0$ toma un valor definido, por ejemplo $f(0) = 45$. La gráfica de la función $Q = f(t)$ está dibujada en la figura 7.

⁴ Esto no significa que tal expresión sea imposible. En el capítulo 12 se mostrará cómo obtener una fórmula única.

Hasta aquí hemos presentado varios ejemplos de funciones expresadas por fórmulas. La posibilidad de representar una función mediante fórmulas es extremadamente importante desde el punto de vista matemático, pues aquéllas proporcionan condiciones muy favorables para investigar las propiedades de las funciones por métodos matemáticos.

Pero no debemos pensar que las fórmulas son el único modo de definir una función. Hay muchos otros métodos; por ejemplo, la gráfica de una función, que da una imagen geométrica de ella. El siguiente ejemplo constituye una buena ilustración de este método.

Para registrar la variación de la temperatura del aire en el curso de veinticuatro horas, las estaciones meteorológicas utilizan un instrumento llamado termógrafo. Un termógrafo consta de un tambor que gira sobre su eje mediante un mecanismo de relojería, y de una pieza de latón curvada y extremadamente sensible a los cambios de temperatura. Cuando la temperatura aumenta sube una plumilla que va sujeta a la pieza metálica por un sistema de palancas; y recíprocamente, una disminución de la temperatura hace bajar a la pluma. Sobre el tambor va enrollada una cinta de papel milimetrado, sobre la cual traza la pluma una línea continua: la gráfica de la función $T = f(t)$, que expresa la dependencia entre el tiempo y la temperatura del aire. De esta gráfica podemos determinar, sin cálculo alguno, el valor de la temperatura en cualquier instante de tiempo t .

Este ejemplo demuestra que una gráfica determina en sí misma una función, independientemente de si la función dada viene expresada por una fórmula o no.

Más adelante volveremos sobre esta cuestión (ver Cap. 12) y demostraremos el siguiente e importante aserto: toda gráfica continua puede representarse por una fórmula o, como también se dice, por una expresión analítica. Esta afirmación es válida también para muchas gráficas discontinuas⁵.

Hagamos notar que la validez de esta afirmación, que es de gran importancia teórica, no fue admitida con carácter general hasta mediados del siglo pasado. Hasta ese momento los matemáticos sólo entendían por el término «función» una expresión analítica (fórmula). Y es que estaban bajo la impresión equivocada de que muchas gráficas discontinuas no correspondían a ninguna expresión analítica, pues suponían que si una función podía expresarse por una fórmula,

⁵ Naturalmente, la afirmación anterior sólo resultará completamente clara al lector después de dar una definición precisa de lo que se entiende en matemáticas por el término «fórmula» y «expresión analítica».

entonces su gráfica debía poseer ciertas propiedades particularmente deseables en comparación con las otras gráficas.

Pero en el siglo XIX se descubrió que toda gráfica continua puede venir representada por una fórmula más o menos complicada. Así se debilitó el papel excepcional de la expresión analítica como medio de definición de funciones y se llegó a la nueva y más flexible definición dada anteriormente. Según esta definición, una variable y se dice función de una variable x si existe una regla por la cual a todo valor de x en el dominio de definición de la función corresponde un valor completamente determinado y , independientemente del modo en que se dé esta regla: mediante una fórmula, una gráfica, una tabla o de cualquier otro modo.

Conviene señalar aquí que en la literatura matemática la definición anterior de función se asocia a menudo con el nombre de Dirichlet, pero esta definición la dio simultánea e independientemente N. Y. Lobachevski. Sugerimos finalmente como ejercicio, que el lector dibuje las gráficas de las funciones x^3 , $\sqrt{x}$, $\sin x$, $\sin 2x$, $\sin(x + \pi/4)$, $\ln x$, $\ln(1 + x)$, $|x - 3|$, $(x + |x|)/2$.

Conviene observar también que la gráfica de una función que satisface la relación

$$f(-x) = f(x)$$

para todos los valores de x , es simétrica con respecto al eje y , y que en el caso

$$f(-x) = -f(x)$$

la gráfica es simétrica respecto al origen de coordenadas. Considérese también cómo obtener la gráfica de una función $f(a + x)$, donde a es una constante, a partir de la gráfica de $f(x)$. Finalmente, el lector debe estudiar cómo se pueden encontrar, mediante las gráficas de las funciones $f(x)$ y $\phi(x)$, los valores de la función $y = f[\phi(x)]$.

§3. Límites

En el apartado §1 se estableció que el análisis matemático moderno utiliza un método especial, que fue elaborado en el transcurso de muchos siglos, y constituye ahora su instrumento básico. Nos referimos al método de los infinitésimos o, lo que en esencia es lo mismo, de los límites. Trataremos de dar alguna idea de estos conceptos. Con este objeto, examinemos el ejemplo siguiente.

Deseamos calcular el área limitada por la parábola de ecuación $y = x^2$, el eje x y la recta $x = 1$ (Fig. 8). La matemática elemental no nos proporciona ningún método para resolver este problema. Pero he aquí cómo podemos proceder.

Dividimos el intervalo $[0, 1]$ a lo largo del eje x en n partes iguales por los puntos

$$0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$$

y sobre cada una de estas partes construimos el rectángulo cuyo lado vertical izquierdo llega hasta la parábola. Como resultado obtenemos el conjunto de rectángulos rayados de la figura 8, la suma S_n de cuyas áreas viene dada por

$$\begin{aligned} S_n &= 0 \cdot \frac{1}{n} + \left(\frac{1}{n}\right)^2 \frac{1}{n} + \left(\frac{2}{n}\right)^2 \frac{1}{n} + \dots + \left(\frac{n-1}{n}\right)^2 \frac{1}{n} = \\ &= \frac{1^2 + 2^2 + \dots + (n-1)^2}{n^3} = \frac{(n-1)n(2n-1)}{6n^3}. \end{aligned} \quad ^6$$

Expresemos S_n en la siguiente forma:

$$S_n = \frac{1}{3} + \left(\frac{1}{6n^2} - \frac{1}{2n}\right) = \frac{1}{3} + \alpha_n. \quad (7)$$

El número α_n , que depende de n , parece a primera vista poco manejable, pero posee una propiedad ciertamente importante: si n aumenta indefinidamente, α_n se aproxima a 0. Esta propiedad puede

⁶ Si en las igualdades evidentes $(k+1)^3 - k^3 = 3k^2 + 3k + 1$, se suman los primeros y segundos miembros por separado para los diferentes valores $k = 1, 2, \dots, n-1$, se obtiene la ecuación

$$n^3 - 1 = 3\sigma_n + \frac{3(n-1)n}{2} + n - 1$$

donde $\sigma_n = 1^2 + 2^2 + \dots + (n-1)^2$. Despejando σ_n en esta ecuación, se obtiene

$$\sigma_n = \frac{(n-1)n(2n-1)}{6}.$$

también expresarse como sigue: dado un número positivo arbitrario ϵ , es posible elegir un entero N suficientemente grande tal que, para todo n mayor que N , el número a_n sea menor que el ϵ dado en valor absoluto ⁷.

La variable a_n es un ejemplo de infinitésimo en el sentido en que esta palabra se utiliza en la matemática moderna.

En la figura 8 vemos que, si aumentamos el número n indefinida-

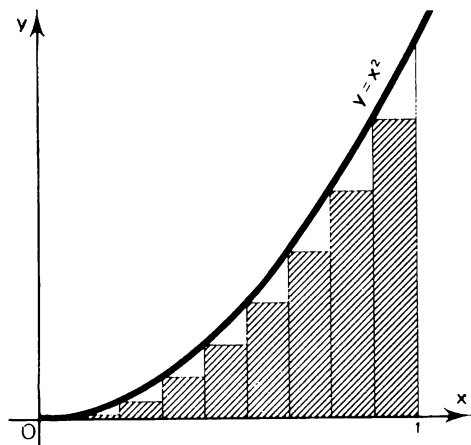


FIG. 8

mente, la suma S_n de las áreas de los rectángulos sombreados se aproximará al área de la figura curvilínea. Por otra parte, y teniendo en cuenta el hecho de que a_n se aproxima a 0 cuando n crece indefinidamente, la ecuación (7) muestra que la suma S_n se aproxima al número $1/3$. De aquí se sigue que el área buscada de la figura S es igual a $1/3$, con lo cual queda resuelto el problema.

Así, este método equivale a lo siguiente: para encontrar una cierta magnitud S , introducimos otra magnitud S_n , *variable*, que se aproxima a S a través de valores particulares $S_1, S_2, S_3, \dots$, que se corresponden, según una cierta ley, con los números naturales $n = 1, 2, \dots$.

⁷ Por ejemplo, si $\epsilon = 0,001$, podemos tomar $N = 500$. En efecto, puesto que

$$\frac{1}{6n^2} < \frac{1}{2n}$$

para n entero positivo, se tiene

$$|a_n| = \left| \frac{1}{6n^2} - \frac{1}{2n} \right| = \frac{1}{2n} - \frac{1}{6n^2} < \frac{1}{2n} < 0,001$$

para todo $n > 500$. Del mismo modo sería posible asignar a ϵ valores arbitrariamente pequeños, por ejemplo:

$$\epsilon_1 = 0,0001, \quad \epsilon_2 = 0,00001, \dots,$$

y para cada uno de ellos elegir, como antes, valores apropiados $N = N_1, N_2, \dots$.

Del hecho de que la variable S_n pueda representarse como la suma de una constante $1/3$ y un infinitésimo α_n , concluimos entonces que S_n se aproxima a $1/3$ y que $S = 1/3$. En el lenguaje de la moderna teoría de límites podemos decir que cuando n crece indefinidamente la magnitud variable S_n se aproxima a un límite, que es igual a $1/3$.

Vamos a dar ahora una definición rigurosa de los conceptos introducidos.

Si una magnitud variable α_n ($n = 1, 2, \dots$) tiene la propiedad de que para todo número positivo ϵ arbitrariamente pequeño es posible elegir un entero N positivo suficientemente grande y tal que para todo $n > N$ se tenga $|\alpha_n| < \epsilon$, entonces decimos que α_n es un *infinitésimo* y escribimos

$$\lim_{n \rightarrow \infty} \alpha_n = 0 \quad \text{ó} \quad \alpha_n \rightarrow 0.$$

Por otra parte, si una variable x_n se puede representar como una suma

$$x_n = a + \alpha_n,$$

donde a es constante y α_n un infinitésimo, entonces decimos que la variable x_n , cuando n crece indefinidamente, se aproxima, o tiende, al número a , y escribimos

$$\lim x_n = a \quad \text{o bien} \quad x_n \rightarrow a.$$

El número a se llama *límite* de x_n . En particular, el límite de un infinitésimo es, evidentemente, cero.

Consideremos los siguientes ejemplos de magnitudes variables

$$x_n = \frac{1}{n}, y_n = -\frac{1}{n^2}, z_n = \frac{(-1)^n}{n}, u_n = \frac{n-1}{n} = 1 - \frac{1}{n},$$

$$v_n = (-1)^n \quad (n = 1, 2, \dots).$$

Es claro que x_n, y_n y z_n son infinitésimos; el primero tiende a cero mediante valores decrecientes; el segundo, por valores negativos crecientes, mientras que el tercero lo hace por valores que oscilan alrededor de cero. Además, $u_n \rightarrow 1$, mientras que v_n no tiene límite, puesto que cuando n crece no se aproxima a ningún número, sino que oscila continuamente entre los valores 1 y -1 .

Otro concepto importante en análisis es el de *infinito* o infinitamente grande, que se define como una variable x_n ($n = 1, 2, \dots$), con la propiedad de que, para todo número positivo M arbitrariamente grande, se puede encontrar un número N tal que para todo $n > N$

$$|x_n| > M.$$

El hecho de que la magnitud x_n sea un infinito se nota así:

$$\lim_{n \rightarrow \infty} x_n = \infty \quad \text{ó} \quad x_n \rightarrow \infty.$$

Una magnitud x_n tal se dice que tiende a infinito. Si es positiva (negativa) de un valor en adelante, se nota así: $x_n \rightarrow +\infty$ ($x_n \rightarrow -\infty$). Por ejemplo, para $n = 1, 2, \dots$,

$$\lim n^2 = +\infty, \quad \lim (-n^3) = -\infty;$$

$$\lim \log \frac{1}{n} = -\infty, \quad \lim \operatorname{tg} \left(\frac{\pi}{2} + \frac{1}{n} \right) = -\infty.$$

Se comprueba fácilmente que, si una magnitud a_n es un infinito, entonces $\beta_n = 1/a_n$ es un infinitésimo, y recíprocamente.

Dos magnitudes variables x_n e y_n pueden ser sumadas, restadas, multiplicadas y divididas la una por la otra, para obtener nuevas magnitudes que son, en general, también variables: su suma $x_n + y_n$, su diferencia $x_n - y_n$, su producto $x_n y_n$, y su cociente x_n/y_n . Sus correspondientes valores particulares serán:

$$x_1 \pm y_1, x_2 \pm y_2, x_3 \pm y_3, \dots$$

$$x_1 y_1, x_2 y_2, x_3 y_3, \dots$$

$$\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{x_3}{y_3}, \dots$$

También es posible demostrar, como se deduce fácilmente, que, si las variables x_n e y_n tienen límites finitos, su suma, diferencia, producto y cociente tienen también límites finitos, que son respectiva-

mente iguales a la suma, diferencia, producto y cociente de estos límites. Esto se puede expresar así:

$$\lim (x_n \pm y_n) = \lim x_n \pm \lim y_n; \lim (x_n y_n) = \lim x_n \lim y_n;$$

$$\lim \frac{x_n}{y_n} = \frac{\lim x_n}{\lim y_n}.$$

Sin embargo, en el caso del cociente es necesario suponer que el límite del denominador ($\lim y_n$) no es cero. Si $\lim y_n = 0$ y $\lim x_n \neq 0$, entonces el cociente de x_n a y_n no tendrá un límite finito, sino que tenderá a infinito.

Especialmente interesante, y al mismo tiempo importante, es el caso en que el numerador y el denominador tienden simultáneamente a cero. Aquí es imposible saber de antemano si el cociente x_n/y_n tendrá límite y, en caso afirmativo, cuál será ese límite, puesto que depende completamente del modo en que x_n e y_n se aproximen a cero. Por ejemplo, si

$$x_n = \frac{1}{n}, y_n = \frac{1}{n^2}, z_n = \frac{(-1)^n}{n} \quad (n = 1, 2, \dots),$$

entonces

$$\frac{y_n}{x_n} = \frac{1}{n} \rightarrow 0, \quad \frac{x_n}{y_n} = n \rightarrow \infty.$$

Por otra parte, el cociente

$$\frac{x_n}{z_n} = (-1)^n$$

evidentemente no tiende a ningún límite.

Así, el caso en el que el numerador y el denominador de la fracción tienden simultáneamente a cero no puede ser estudiado mediante teoremas generales, y es necesario hacer un estudio especial para cada fracción particular de esta clase.

Veremos más tarde que el problema fundamental del cálculo diferencial —que puede ser considerado como el problema de determinar la velocidad de un movimiento no uniforme en un instante dado— se reduce a determinar el límite del cociente de dos infinitésimos: el incremento de la distancia recorrida por el incremento del tiempo.

Hasta ahora hemos considerado variables x_n que toman una sucesión de valores numéricos $x_1, x_2, x_3, \dots, x_n, \dots$, variando el índice n a lo largo de la sucesión de los números naturales $n = 1, 2, 3, \dots$. Pero también se puede considerar el caso de que n varíe continuamente (como el tiempo, por ejemplo) y determinar también el límite de la variable x_n . Las propiedades de tales límites son del todo análogas a las formuladas anteriormente para variables discretas (esto es, discontinuas). Nótese que el hecho de que n crezca indefinidamente es irrelevante. Igualmente posible sería considerar el caso en que, variando continuamente, n se aproximase a un valor dado n_0 .

Como ejemplo estudiemos la variación de $(\text{sen } x)/x$ cuando x se aproxima a cero. La tabla 1 muestra valores de este cociente para ciertos valores de x :

TABLA 1

x	$\frac{\text{sen } x}{x}$
0,50	0,9589 ...
0,10	0,9983 ...
0,05	0,9996 ...
...	...

(se supone que los valores de x están medidos en radianes).

Es obvio que cuando x tiende (o se aproxima) a cero, el valor de $(\text{sen } x)/x$ tiende a uno, pero naturalmente debemos dar una demostración rigurosa de este hecho. La demostración puede obtenerse, por ejemplo, a partir de la siguiente desigualdad, válida para todos los ángulos no nulos del primer cuadrante:

$$\text{sen } x < x < \text{tg } x.$$

Si dividimos los términos de esta desigualdad por $\text{sen } x$, obtenemos

$$1 < \frac{x}{\text{sen } x} < \frac{1}{\cos x},$$

de donde se deduce que

$$\cos x < \frac{\text{sen } x}{x} < 1.$$

Pero cuando x tiende a cero a través de valores positivos, $\cos x$ se aproxima a 1, por lo que el valor de $(\sin x)/x$, contenido en el intervalo entre $\cos x$ y 1, también se aproxima a 1; esto es:

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$$

Más tarde tendremos ocasión de hacer uso de este resultado.

La expresión anterior fue obtenida para el caso de que x tienda a cero tomando valores positivos. Pero modificando la demostración de un modo obvio, se puede obtener el mismo resultado cuando x tiende a cero tomando valores negativos.

Analicemos ahora la siguiente cuestión. Una magnitud variable puede tener o no límite, y surge la pregunta de si es posible dar un criterio para determinar la existencia del límite de una variable. Nos limitaremos a un caso importante y suficientemente general, para el que puede darse un tal criterio. Supongamos que la magnitud variable x_n crece, o que al menos no decrece; esto es, satisface las desigualdades

$$x_1 \leq x_2 \leq x_3 \leq \dots,$$

y supongamos también que ninguno de sus valores supera a cierto número fijo M ; es decir, $x_n \leq M$ ($n = 1, 2, \dots$). Si situamos los valores de x_n y el número M sobre el eje x , vemos que el punto variable x_n se mueve a lo largo del eje x hacia la derecha, pero permaneciendo siempre a la izquierda del punto M . Es bastante evidente que el punto variable x_n debe aproximarse a un cierto punto límite a , situado a la izquierda de M o, a lo sumo, coincidente con M .

Así, en el caso que estamos considerando, existe el límite

$$\lim x_n = a$$

de la variable x_n .

El razonamiento anterior es de carácter intuitivo, pero podemos considerarlo como una demostración. En un curso de análisis moderno se da una demostración rigurosa de este hecho basada en la teoría de los números reales.

Como ejemplo consideremos la variable

$$u_n = \left(1 + \frac{1}{n}\right)^n \quad (n = 1, 2, 3, \dots).$$

Los primeros valores son $u_1 = 2$, $u_2 = 2,25$, $u_3 \approx 2,37$, $u_4 \approx 2,44$, $\dots$, que se ve van creciendo. Por el teorema del binomio de Newton se puede demostrar que este crecimiento se mantiene para todo n . Además es fácil demostrar también que la desigualdad $u_n < 3$ es válida para todo n . De este modo, la variable debe tener un límite no superior a 3. Veremos que este límite juega un papel muy importante en la física y la matemática y en cierto sentido es la base más natural para los logaritmos.

Este límite, que se suele notar por la letra e , es igual a

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = 2,718281828459045\dots$$

Un análisis más detallado demuestra que el número e no es racional ⁸.

Es posible demostrar también que el límite que estamos considerando existe y es igual a e , no sólo cuando $n \rightarrow +\infty$, sino también cuando $n \rightarrow -\infty$. En ambos casos n puede tomar valores no enteros.

Mencionemos una importante aplicación del concepto límite a la física. Consiste en el hecho de que sólo utilizando el concepto de límite (paso al límite) es posible dar una definición rigurosa de muchas magnitudes físicas.

Consideremos también por un momento el siguiente ejemplo geométrico. En geometría elemental las figuras que primero se consideraron fueron las limitadas por segmentos de rectas. Pero más tarde se planteó la tarea de encontrar la longitud de una circunferencia de radio dado. Si analizamos las dificultades relacionadas con la solución de este problema encontramos que se reducen a lo siguiente.

Hay que resolver la cuestión de qué es lo que se entiende por longitud de una circunferencia; esto es, debemos dar una definición rigurosa de esta longitud. Es esencial que la definición venga expresada en términos de longitudes de segmentos rectilíneos y que nos proporcione la posibilidad del cálculo efectivo de la longitud de la circunferencia.

Se comprende, naturalmente, que el resultado de este cálculo deberá estar de acuerdo con la experiencia diaria. Por ejemplo, si

⁸ En conexión con esto hay que observar que la adición, sustracción, multiplicación y división (excluida la división por cero) de números racionales (esto es, de números de la forma p/q donde p y q son enteros) conduce a números racionales. Pero este no es necesariamente el caso para la operación de paso al límite. El límite de una sucesión de números racionales puede ser un número irracional.

consideramos una circunferencia construida con un hilo, al cortarlo y estirarlo se debe obtener un segmento cuya longitud, dentro de los límites de precisión de la medida, coincida con la longitud calculada.

Como se sabe por geometría elemental, la solución de este problema se obtiene con la definición siguiente: la longitud de una circunferencia es el límite a que tiende el perímetro de un polígono regular⁹ inscrito en ella cuando el número de lados del polígono tiende a infinito. Así la solución del problema se apoya esencialmente en el concepto de límite.

La longitud de una curva suficientemente regular se define del mismo modo. En las secciones que siguen encontraremos cierto número de ejemplos de magnitudes físicas y geométricas que únicamente se pueden definir mediante el concepto de límite.

Los conceptos de límite e infinitésimo recibieron una formulación definitiva a principios del siglo pasado. Las definiciones introducidas hasta aquí están relacionadas con el nombre de Cauchy; los matemáticos anteriores a él operaban con conceptos mucho menos precisos. Los conceptos actuales de límite, de infinitésimo como magnitud variable y de número real han resultado del desarrollo del análisis matemático, y sirvieron al mismo tiempo como medio para plantear y clarificar sus muchas conquistas.

§4. *Funciones continuas*

Las funciones continuas constituyen la clase básica de funciones para las operaciones del análisis matemático. La idea general de función continua viene a ser la de que su gráfica sea continua; esto es, que la curva pueda dibujarse sin separar el lápiz del papel.

Una función continua proporciona la expresión matemática de una situación que aparece con frecuencia en la vida diaria: el hecho de que a un pequeño crecimiento de la variable independiente corresponda un pequeño incremento en la variable dependiente, o función. Ejemplos importantes de funciones continuas los proporcionan las diversas leyes que rigen el movimiento de los cuerpos, $s = f(t)$, y que expresan la dependencia de la distancia s respecto del tiempo t . Puesto que el tiempo y la distancia son continuos, una ecuación del movi-

⁹ No es importante que el polígono sea regular. Lo único esencial es que el lado mayor del polígono inscrito variable tienda a cero.

miento de un cuerpo $s = f(t)$ define entre ellos una relación continua, caracterizada por el hecho de que un pequeño incremento en el tiempo se traduce en un incremento también pequeño de la distancia.

El hombre llegó al concepto abstracto de continuidad observando los medios densos que le rodean, tanto sólidos, líquidos como gaseosos; por ejemplo, los metales, el agua y el aire. En realidad, como ahora sabemos, todo medio físico representa la acumulación de un gran número de partículas distintas en movimiento. Pero estas partículas y sus distancias mutuas son tan pequeñas en comparación con las dimensiones del medio en el que tienen lugar los fenómenos de la física macroscópica, que muchos de estos medios se suponen aproximadamente carentes de intersticios, esto es, como distribuciones continuas en el espacio que ocupan. En esta suposición se basan muchas de las ciencias físicas, por ejemplo la hidrodinámica, la aerodinámica y la teoría de la elasticidad. El concepto matemático de continuidad juega un gran papel en estas ciencias, así como en muchas otras.

Consideremos una función arbitraria $y = f(x)$ y un valor particular de la variable independiente x_0 . Si la función representa un proceso continuo, entonces a valores x que difieran sólo ligeramente de x_0 corresponderán valores de la función $f(x)$ que difieren sólo ligeramente del valor $f(x_0)$ en el punto x_0 . Así, si el incremento $x - x_0$ de la variable independiente es pequeño, el incremento correspondiente $f(x) - f(x_0)$ de la función será también pequeño. En otras palabras, si el incremento de la variable independiente $x - x_0$ tiende a cero, entonces el incremento $f(x) - f(x_0)$ de la función debe aproximarse a cero, hecho que puede expresarse del modo siguiente:

$$\lim_{(x-x_0) \rightarrow 0} [f(x) - f(x_0)] = 0. \quad (8)$$

Esta relación constituye la definición matemática de continuidad de la función en el punto x_0 ; es decir, la función $f(x)$ se dice *continua en el punto* x_0 si se verifica la igualdad (8).

Damos finalmente la siguiente definición. Una función se dice *continua en un intervalo dado* si es continua en todo punto x_0 de este intervalo; esto es, si se verifica la igualdad (8) en cada uno de sus puntos.

Así, para dar una definición matemática de esa propiedad de las funciones que viene caracterizada por el hecho de que su gráfica

sea continua (en el sentido usual de la palabra), fue necesario definir primero la continuidad local (continuidad en el punto x_0), y luego, a partir de ella, definir la continuidad de la función en todo el intervalo.

Esta definición, introducida por primera vez a comienzos del siglo pasado por Cauchy, es la que se adopta ahora en general en el análisis matemático contemporáneo. Su empleo en muchos ejemplos concretos ha demostrado que se corresponde muy bien con la noción intuitiva de función continua, cual es, por ejemplo, la de venir representada por una gráfica continua.

Como ejemplos de funciones continuas puede considerar el lector las conocidas funciones elementales x^n , $\operatorname{sen} x$, $\cos x$, a^x , $\log x$, $\operatorname{arc} \operatorname{sen} x$, $\operatorname{arc} \cos x$. Todas estas funciones son continuas en los intervalos en los que están definidas.

Si se suman, restan, multiplican o dividen (excepto la división por cero) funciones continuas, el resultado es también una función continua. Pero en el caso de la división la continuidad desaparece en general en aquellos valores x_0 para los cuales se anula la función del denominador. El resultado de la división es en este caso una función que es *discontinua* en el punto x_0 .

La función $y = 1/x$ puede servir como ejemplo de función que es discontinua en el punto $x = 0$. Otras funciones discontinuas vienen representadas por las gráficas de la figura 9.

Recomendamos al lector que examine estas gráficas cuidadosamente. Observará que las discontinuidades de las funciones representadas son de diferentes clases: en algunos casos existe el límite de $f(x)$ cuando x tiende al punto x_0 (donde la función es discontinua), pero este límite es diferente de $f(x_0)$. En otros casos, como en la figura 9c, el límite sencillamente no existe. Puede suceder también que cuando x tiende a x_0 por la derecha $f(x) - f(x_0) \rightarrow 0$, pero cuando $x \rightarrow x_0$ por la izquierda, $f(x) - f(x_0)$ no tienda a cero, o recíprocamente. En estos casos, naturalmente, la función tiene una discontinuidad, pero podemos decir que en ese punto es «continua lateralmente», «por la izquierda» o «por la derecha». Todos estos casos están representados en las gráficas de la figura 9.

Como ejercicio recomendamos al lector que considere los valores que hay que dar a las funciones

$$\frac{\operatorname{sen} x}{x}, \frac{1 - \cos x}{x^2}, \frac{x^3 - 1}{x - 1}, \frac{\operatorname{tg} x}{x}$$

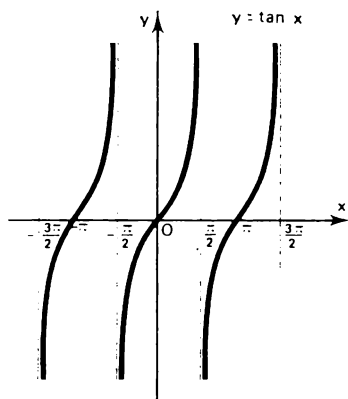


FIG. 9a

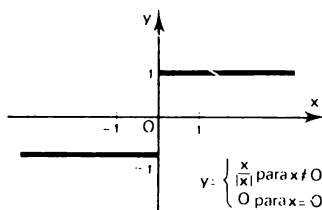


FIG. 9b

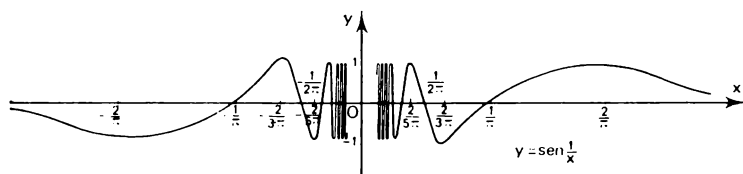


FIG. 9c

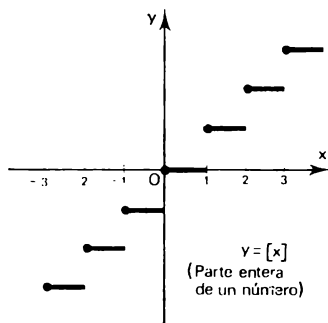


FIG. 9d

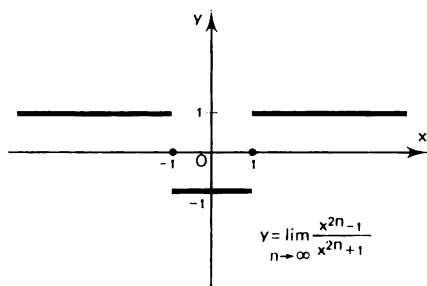


FIG. 9e

en los puntos en los que no estén definidas (esto es, en los puntos en los que el denominador es cero), para que puedan ser continuas en estos puntos. ¿Es posible encontrar tales números para las funciones

$$\operatorname{tg} x, \frac{1}{x-1}, \frac{x-2}{x^2-4} ?$$

Las funciones discontinuas representan en la matemática los numerosos procesos de salto que se observan en la naturaleza. Cuando un cuerpo se ve impelido súbitamente por un golpe de viento, por ejemplo, el valor de la velocidad varía bruscamente. Las transiciones cualitativas se producen muchas veces a saltos. En el §2 introdujimos la función $Q = f(t)$, que expresaba el modo en el que la cantidad de calor en una masa de agua (o hielo) depende de la temperatura. En el entorno del punto de fusión del hielo la cantidad de calor $Q = f(t)$ experimenta un salto cuando varía t .

Las funciones con discontinuidades aisladas son muy corrientes en análisis. Pero como ejemplo de función más complicada, donde el número de discontinuidades es infinito, consideremos la llamada función de Riemann, que es igual a cero en todos los puntos irracionales e igual a $1/q$ en los puntos racionales de la forma $x = p/q$ (donde p/q es una fracción irreducible). Esta función es discontinua en todos los puntos racionales y continua en los puntos irracionales. Pero alterándola ligeramente puede obtenerse con facilidad un ejemplo de función que es discontinua en todos los puntos¹⁰. Observemos de paso que incluso para funciones tan complicadas el análisis moderno ha descubierto muchas propiedades interesantes, que se investigan en una de las ramas del análisis, la teoría de funciones de variable real. Esta teoría se ha desarrollado con extraordinaria rapidez en los últimos cincuenta años.

§5. Derivada

El siguiente concepto fundamental del análisis es el de *derivada*. Consideremos los dos problemas de donde surgió históricamente.

Velocidad

Al comienzo del presente capítulo definimos la velocidad de caída de un cuerpo. Para ello hicimos uso del paso al límite de una velocidad media calculada para distancias finitas a la velocidad en

¹⁰ Basta hacer la función igual a 1 en los puntos irracionales.

un punto en un instante dado. El mismo procedimiento puede utilizarse para definir la velocidad instantánea de un movimiento no uniforme cualquiera. De hecho, sea

$$s = f(t) \quad (9)$$

una función que expresa la dependencia de la distancia s , recorrida por el punto material, respecto del tiempo t . Para encontrar la velocidad en el instante $t = t_0$ consideremos el intervalo de tiempo de t_0 a $t_0 + h$ ($h \neq 0$). En este intervalo el punto recorrerá la distancia

$$\Delta s = f(t_0 + h) - f(t_0).$$

La velocidad media v_m en esta parte del recorrido dependerá de h :

$$v_m = \frac{\Delta s}{h} = \frac{1}{h} [f(t_0 + h) - f(t_0)],$$

que representará la velocidad verdadera en el punto t_0 con mayor precisión a medida que h sea cada vez más pequeño. De aquí se sigue que la velocidad verdadera en el tiempo t_0 es igual al límite

$$v = \lim_{h \rightarrow 0} \frac{f(t_0 + h) - f(t_0)}{h}$$

del cociente entre el incremento en la distancia y el incremento en el tiempo cuando el último tiende a cero (aunque sin llegar realmente a anularse). Para calcular la velocidad de diferentes clases de movimiento, debemos encontrar un procedimiento de calcular este límite para diferentes funciones $f(t)$.

Tangente

Un límite análogo al anterior aparece también en conexión con otro problema, esta vez geométrico: el de dibujar una tangente a una curva plana cualquiera.

Se la curva C la gráfica de una función $y = f(x)$, y sea A el punto que, perteneciendo a la curva C , tiene como abscisa x_0 (Fig. 10). ¿A qué línea recta llamaremos tangente a C en el punto A ? En geometría elemental esta cuestión no se plantea. La única curva que

se estudia a ese nivel, la circunferencia, permite definir la tangente como una línea recta que sólo tiene un punto en común con la curva.

Pero para otras curvas dicha definición no se corresponde claramente con nuestro concepto de «tangencia». Así, de las dos rectas L y M de la figura 11, la primera evidentemente no es tangente a la curva dibujada (una curva sinusoidal), aunque sólo tiene un punto en común con ella; mientras que la segunda recta tiene muchos puntos en común con la curva y sin embargo es tangente a la curva en todos ellos.

Para definir la tangente, consideremos sobre la curva C (Fig. 10) otro punto A' , distinto de A , de abscisa $x_0 + h$. Dibujemos la secante

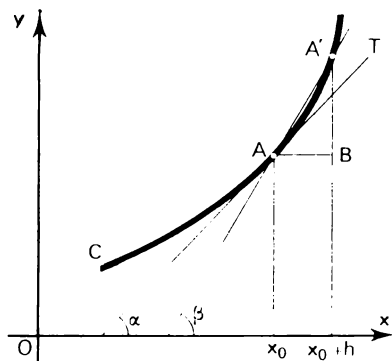


FIG. 10

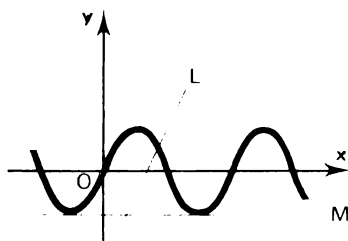


FIG. 11

AA' y notemos por β el ángulo que forma con el eje x . Dejemos ahora que el punto A' se aproxime al A a lo largo de la curva C . Si la secante correspondiente AA' tiende a una posición límite, entonces la recta T que corresponde a esta posición límite se llama *tangente* en el punto A . Evidentemente, el ángulo α formado por la recta T con el eje x debe ser igual al valor límite del ángulo variable β .

El valor de $\operatorname{tg} \beta$ se determina fácilmente en el triángulo ABA' (Fig. 10):

$$\operatorname{tg} \beta = \frac{BA'}{AB} = \frac{f(x_0 + h) - f(x_0)}{h}.$$

Para la posición límite debemos tener

$$\operatorname{tg} \alpha = \lim_{A' \rightarrow A} \operatorname{tg} \beta = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

esto es, la tangente trigonométrica del ángulo de inclinación de la recta tangente es igual al límite del cociente entre el incremento de la función $f(x)$ en el punto x_0 y el incremento correspondiente en la variable independiente, cuando este último tiende a cero sin llegar a anularse.

Veamos otro ejemplo más, que conduce al cálculo de un límite análogo. Supongamos que una corriente eléctrica variable atraviesa un conductor y que conocemos la función $Q = f(t)$, que expresa la cantidad de electricidad que ha pasado por una sección transversal dada del conductor hasta el tiempo t . En el intervalo de t_0 a $t_0 + h$ pasará por esta sección transversal una cantidad de electricidad ΔQ igual a $f(t_0 + h) - f(t_0)$. El valor medio de la corriente será, por tanto, igual a

$$I_m = \frac{\Delta Q}{h} = \frac{f(t_0 + h) - f(t_0)}{h}.$$

El límite de este cociente cuando $h \rightarrow 0$ dará el valor de la corriente en el tiempo t_0

$$I = \lim_{h \rightarrow 0} \frac{f(t_0 + h) - f(t_0)}{h}.$$

A pesar de que se refieren a tres diferentes ramas de la ciencia —mecánica, geometría y electricidad—, los tres problemas planteados conducen a una misma operación matemática a realizar sobre una función dada: la operación de encontrar el límite del cociente entre el incremento de la función y el incremento correspondiente h de la variable independiente cuando $h \rightarrow 0$. El número de problemas de este tipo podría aumentarse indefinidamente, y sus soluciones conducirían a la misma operación. A ella conduce, por ejemplo, el estudio de la velocidad de una reacción química, o de la densidad de una masa no homogénea, etc. En vista del excepcional papel que desempeña esta operación, ha recibido un nombre especial, derivación, y el resultado de la operación se llama *derivada* de la función.

Así, la *derivada de la función* $y = f(x)$, o con más precisión el

valor de la derivada en un punto dado x , es el límite¹¹ del cociente entre el incremento $f(x+h) - f(x)$ de la función y el incremento h de la variable independiente cuando h tiende a cero. Se suele escribir $h = \Delta x$, y $f(x + \Delta x) - f(x) = \Delta y$; en ese caso la definición de derivada se escribe en la forma concisa:

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}.$$

El valor de la derivada depende evidentemente del punto x en el cual se calcula. Así la derivada de una función $y = f(x)$ es a su vez una función de x . Se acostumbra a notar la derivada así:

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}.$$

Se utilizan también frecuentemente otras notaciones para la derivada:

$$\frac{df(x)}{dx}, \frac{dy}{dx}, y', y'_x.$$

Hay que tener en cuenta que aunque la notación dy/dx se parece a la de las fracciones, se lee como un símbolo para la derivada. En las secciones siguientes el numerador y el denominador de esta «fracción» adquirirán un sentido independiente, de modo tal que su cociente coincidirá con la derivada; esto justifica la notación anterior.

Los resultados de estos ejemplos pueden formularse ahora como sigue.

La velocidad de un punto para el cual la distancia s es una función dada del tiempo t , $s = f(t)$, es igual a la derivada de esta función

$$v = s' = f'(t).$$

Es decir, la velocidad es la derivada de la distancia respecto al tiempo.

La tangente trigonométrica del ángulo de inclinación de la recta tangente a la curva $y = f(x)$ en el punto de abscisa x es igual a la derivada de la función $f(x)$ en este punto:

$$\operatorname{tg} \alpha = y' = f'(x).$$

¹¹ Se sobreentiende que hablamos aquí del caso en que realmente existe el límite en cuestión. Si este límite no existe, se dice que en el punto x la función no es derivable.

La intensidad de la corriente I en el tiempo t (siendo $Q = f(t)$ la cantidad de electricidad que ha pasado por una sección transversal del conductor en el tiempo t) es igual a la derivada:

$$I = Q' = f'(t).$$

Hay que tener en cuenta lo siguiente. La velocidad de un movimiento no uniforme en un instante dado es un concepto puramente físico, que surge de la experiencia. El hombre llegó a él como resultado de numerosas observaciones sobre movimientos concretos diferentes. El estudio del movimiento no uniforme de un cuerpo en distintas partes de su trayectoria, la comparación de distintos movimientos de esta clase que tienen lugar simultáneamente y, en particular, el estudio de los fenómenos de colisión de los cuerpos representan todos una acumulación de experiencias prácticas que condujeron a la aparición del concepto físico de velocidad de un movimiento no uniforme en un instante dado. Pero la definición rigurosa de velocidad dependía necesariamente del modo de definir su valor numérico, y esto sólo fue posible con el concepto de derivada.

En mecánica, la velocidad, en un instante t , de un cuerpo que se mueve de acuerdo con la ecuación $s = f(t)$ se define como la derivada de la función $f(t)$ para este valor de t .

Lo dicho al comienzo de la presente sección demuestra, por una parte, la ventaja de introducir la operación de derivación, y por otra, proporciona una justificación razonable de la definición dada anteriormente de velocidad en un instante dado. Así, cuando surgió la cuestión de obtener la velocidad de un punto con movimiento no uniforme sólo teníamos en realidad una noción empírica de su valor, pero no una definición exacta. Ahora, como resultado de nuestro análisis, hemos obtenido una definición rigurosa del valor de la velocidad en un instante dado: la derivada de la distancia con respecto al tiempo. Este resultado es extremadamente importante desde un punto de vista práctico, puesto que nuestro conocimiento empírico de la velocidad se ha enriquecido notablemente con la posibilidad de realizar cálculos numéricos exactos.

Esto es válido tanto para la intensidad de una corriente, como para otros muchos conceptos que expresan la velocidad de algún proceso, físico, químico, etc.

Esta situación puede servir de ejemplo para otras muchas de naturaleza similar, en las que la experiencia ha llevado a la formación de un concepto relativo al mundo exterior (velocidad, trabajo, densidad, área, etc.) y luego la matemática ha permitido definir este concepto

con precisión, de modo que podamos utilizarlo en los cálculos prácticos.

Ya hemos hecho notar al comienzo del capítulo que el concepto de derivada surgió principalmente como resultado de muchos siglos de esfuerzos dirigidos a resolver dos problemas: dibujar una tangente a una curva y encontrar la velocidad de un movimiento no uniforme. Estos problemas, así como el cálculo de áreas que se discutirá más tarde, interesaron ya a los matemáticos de los tiempos antiguos. Pero hasta el siglo xvi la exposición y el método de resolución de cada problema de esta clase tenían un carácter extremadamente específico. La acumulación de todo este extenso material se vio reducida a un sistema teóricamente completo en el siglo xvii gracias a los trabajos de Newton y Leibnitz. También Euler contribuyó notablemente a la fundamentación del análisis actual.

Pero conviene decir que Newton, Leibnitz y sus contemporáneos suministraron una base lógica muy precaria para sus grandes descubrimientos matemáticos: en sus métodos de razonamiento y en los conceptos con los que operaron había cosas muy poco claras desde nuestro punto de vista. Incluso ya entonces los mismos matemáticos eran conscientes de ello, como lo demuestran las agrias discusiones que se han encontrado en sus correspondencias mutuas. Sin embargo, estos matemáticos de los siglos xvii y xviii realizaron sus actividades puramente matemáticas en íntimo contacto con los estudios de otros investigadores de distintas ramas de las ciencias naturales (física, mecánica, química, tecnología). El planteamiento de un problema matemático se debía generalmente a necesidades prácticas o al deseo de entender algún fenómeno de la naturaleza, y tan pronto como el problema era resuelto, la solución se sometía de un modo u otro a pruebas experimentales. Por consiguiente, a pesar de cierta falta de base lógica, la matemática pudo avanzar en direcciones extremadamente útiles.

Ejemplos de cálculo de derivadas

La definición de derivada como límite

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

permite calcular la derivada de cualquier función particular.

Naturalmente, habrá casos en los que la función no tenga deri-

vada en un punto, o incluso en muchos puntos; en otras palabras, puntos en los que el cociente

$$\frac{f(x+h) - f(x)}{h} \quad (10)$$

cuando $h \rightarrow 0$ no tienda a un límite finito. Esta situación aparece evidentemente en todo punto de discontinuidad de la función $f(x)$, puesto que entonces el cociente (10) tiene un numerador que no tiende a cero mientras el denominador sí. En ocasiones la derivada tampoco existe en puntos donde la función es continua. Un sencillo

ejemplo lo dan aquellos puntos de la gráfica donde la función forma un ángulo (Fig. 12). En dicho punto la curva de la gráfica no tiene tangente definida, y por consiguiente la función carece de derivada. A menudo ocurre que en tales puntos la expresión (10) se aproxima a valores diferentes, según h tienda a cero por la derecha o por la izquierda; por lo que, si h tiende a cero de modo arbitrario, el co-

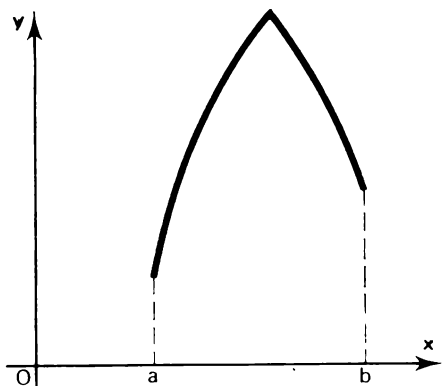


FIG. 12

ciente (0) no tiene límite. Un ejemplo de función más complicada sin derivada viene dado por

$$y = \begin{cases} x \operatorname{sen} \frac{1}{x} & \text{para } x \neq 0, \\ 0 & \text{para } x = 0. \end{cases}$$

La gráfica de esta función está dibujada en la figura 13. En el punto $x = 0$ no tiene derivada porque, como se deduce evidentemente de la gráfica, la secante OA no tiende a ninguna posición definida, incluso cuando $A \rightarrow 0$ lateralmente. De hecho, la secante OA oscila

entre la recta OM y la recta OL . El cociente correspondiente (10) en este caso no tiene límite, incluso si h conserva un mismo signo al tender a cero.

Observemos finalmente que es posible definir, de un modo puramente analítico por medio de una fórmula, una función continua que no tenga derivada en ningún punto. Un ejemplo de tal función fue dado por primera vez por el relevante matemático alemán del siglo pasado Weierstrass.

Por tanto, la clase de las funciones derivables es considerablemente más reducida que la de las funciones continuas.

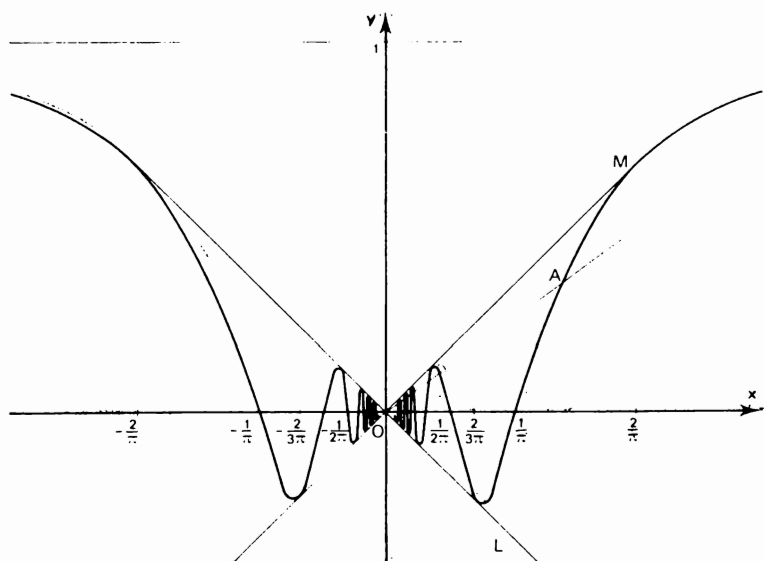


FIG. 13

Pasemos ahora al cálculo de las derivadas de las funciones más sencillas.

1. $y = c$, donde c es una constante. Una constante puede considerarse como un caso especial de función que toma siempre un mismo valor para todo x . Su gráfica es una recta paralela al eje x a una distancia igual a c . La recta forma con el eje x un ángulo $\alpha = 0$, y evidentemente la derivada de una constante es idénticamente

igual a cero: $y' = (c)' = 0$. Desde el punto de vista de la mecánica, esta ecuación dice que la velocidad de un punto en reposo es igual a cero.

2. $y = x^2$

$$\frac{f(x+h) - f(x)}{h} = \frac{(x+h)^2 - x^2}{h} = 2x + h.$$

Cuando $h \rightarrow 0$ obtenemos¹² como límite $2x$; por tanto:

$$y' = (x^2)' = 2x.$$

3. $y = x^n$ (n entero positivo).

$$\begin{aligned} \frac{f(x+h) - f(x)}{h} &= \frac{(x+h)^n - x^n}{h} = \\ &= \frac{x^n + nx^{n-1}h + \frac{n(n-1)}{2!}x^{n-2}h^2 + \dots + h^n - x^n}{h} = \\ &= nx^{n-1} + \frac{n(n-1)}{2!}x^{n-2}h + \dots + h^{n-1}. \end{aligned}$$

Todos los sumandos del miembro de la derecha, a partir del segundo, tienden a cero cuando $h \rightarrow 0$; por consiguiente:

$$y' = (x^n)' = nx^{n-1}.$$

Esta fórmula es válida para cualquier n positivo o negativo, fraccionario e incluso irracional, aunque la demostración debe ser entonces diferente. Utilizaremos este resultado sin demostrarlo. Así, por ejemplo:

$$(\sqrt{x})' = (x^{\frac{1}{2}})' = \frac{1}{2} x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}}, (x > 0);$$

$$(\sqrt[3]{x})' = (x^{\frac{1}{3}})' = \frac{1}{3} x^{-\frac{2}{3}} = \frac{1}{3\sqrt[3]{x^2}}, (x \neq 0);$$

¹² Se supondrá en todos los casos que $h \neq 0$.

$$\left(\frac{1}{x}\right)' = (x^{-1})' = -1 \cdot x^{-2} = -\frac{1}{x^2}, (x \neq 0);$$

$$(x^\pi)' = \pi x^{\pi-1}, (x > 0).$$

$$4. \quad y = \operatorname{sen} x.$$

$$\frac{\operatorname{sen}(x+h) - \operatorname{sen} x}{h} = \frac{2 \operatorname{sen} h/2 \cos(x+h/2)}{h} = \frac{\operatorname{sen} h/2}{h/2} \cdot \cos\left(x + \frac{h}{2}\right).$$

Como ya se explicó, la primera fracción tiende a uno cuando $h \rightarrow 0$, y $\cos(x+h/2)$ se aproxima evidentemente a $\cos x$. Así, la derivada del seno es igual al coseno:

$$y' = (\operatorname{sen} x)' = \cos x.$$

Sugerimos al lector que con la misma clase de razonamiento demuestre que

$$(\cos x)' = -\operatorname{sen} x.$$

5. Ya hicimos notar (capítulo 2, §3) la existencia del límite

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e = 2,71828 \dots$$

Obsérvese también que para el cálculo de este límite no tenía especial importancia el hecho de que n tomase únicamente valores enteros positivos. Sólo depende de que el infinitésimo $1/n$, que se suma a la unidad, y el exponente n , que aumenta indefinidamente, sean recíprocos uno del otro.

Utilizando este resultado, podemos encontrar fácilmente la derivada del logaritmo $y = \log_a x$

$$\frac{\log_a(x+h) - \log_a x}{h} = \frac{1}{h} \log_a \frac{x+h}{x} = \frac{1}{x} \log_a \left(1 + \frac{h}{x}\right)^{x/h}.$$

La continuidad del logaritmo nos permite reemplazar la expresión que se encuentra detrás del signo log por su límite, que es igual a e ; así:

$$\lim_{x/h \rightarrow 0} \left(1 + \frac{h}{x}\right)^{x/h} = e$$

(en este caso el papel de $n \rightarrow \infty$ lo juega el número creciente x/h). Como resultado, obtenemos la regla para derivar un logaritmo

$$(\log_a x)' = \frac{1}{x} \log_a e.$$

Esta regla resulta particularmente sencilla si se elige como base de los logaritmos el número e . Los logaritmos tomados en esta base se llaman *logaritmos naturales* y se notan por $\ln x$. Podemos escribir

$$(\log_e x)' = \frac{1}{x}$$

o también

$$(\ln x)' = \frac{1}{x}.$$

§6. Reglas para la derivación

Los ejemplos estudiados pueden dar la impresión de que el cálculo de la derivada de toda nueva función necesita la invención de nuevos métodos. Este no es el caso. El desarrollo del análisis fue posible, en no poca medida, gracias al descubrimiento de un método unificado y sencillo para obtener la derivada de cualquier función «elemental» (esto es, una función que puede expresarse por una fórmula que consta de una combinación finita de las operaciones algebraicas fundamentales, las funciones trigonométricas, la operación de elevar a una potencia y tomar logaritmos). En la base de este método están las llamadas *reglas de derivación*, consistentes en un cierto número de teoremas que permiten reducir problemas complicados a otros más sencillos.

Explicaremos ahora las reglas de derivación y trataremos de ser breves al deducirlas. Si el lector desea simplemente formarse una idea general del análisis, puede omitir la presente sección, recordando sólo que existen medios de obtener explícitamente la derivada de cualquier función elemental. En este caso será necesario, naturalmente, que dé por válidos los cálculos en posteriores ejemplos.

Derivada de una suma

Supongamos dada y como función de x por la expresión

$$y = \phi(x) + \psi(x),$$

donde $u = \phi(x)$ y $v = \psi(x)$ son funciones conocidas de x . Supongamos además que podemos calcular las derivadas de las funciones u y v . ¿Cómo obtendremos entonces la derivada de la función y ? La solución es sencilla:

$$y' = (u + v)' = u' + v'. \quad (11)$$

En efecto: demos a x un incremento Δx ; u , v e y se incrementarán en Δu , Δv e Δy , relacionadas por la ecuación

$$\Delta y = \Delta u + \Delta v.$$

Así ¹³

$$\frac{\Delta y}{\Delta x} = \frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta x},$$

y después de pasar al límite, para $\Delta x \rightarrow 0$ obtenemos inmediatamente la fórmula (11), suponiendo, naturalmente, que las funciones u y v tengan derivada.

Análogamente podemos deducir la fórmula para derivar la diferencia de dos funciones

$$(u - v)' = u' - v'. \quad (12)$$

Derivada de un producto

La regla para la derivación de un producto es algo más complicada. La derivada del producto de dos funciones, siendo ambas derivables, existe y es igual a la suma del producto de la primera función por la derivada de la segunda más el producto de la segunda por la derivada de la primera; esto es

$$(uv)' = uv' + vu'. \quad (13)$$

En efecto, demos a x un incremento Δx . Entonces las funciones u , v e $y = uv$ se incrementarán respectivamente en Δu , Δv , Δy , satisfaciendo la relación

$$\Delta y = (u + \Delta u)(v + \Delta v) - uv = u \Delta v + v \Delta u + \Delta u \Delta v,$$

de la cual

$$\frac{\Delta y}{\Delta x} = u \frac{\Delta v}{\Delta x} + v \frac{\Delta u}{\Delta x} + \Delta u \frac{\Delta v}{\Delta x}.$$

¹³ Δx no es nunca igual a cero.

Después de pasar al límite para $\Delta x \rightarrow 0$, los dos primeros sumandos del segundo miembro dan el segundo miembro de la fórmula (13), mientras que el tercer sumando se anula¹⁴. Por consiguiente, en el límite obtenemos la regla (13).

En el caso particular $v = c = \text{constante}$, tenemos

$$(cu)' = cu' + uc' = cu', \quad (14)$$

puesto que la derivada de una constante es igual a cero.

Derivada de un cociente

Sea $y = u/v$, donde u y v tienen derivada para un x dado, con $v \neq 0$ para ese valor de x . Evidentemente

$$\Delta y = \frac{u + \Delta u}{v + \Delta v} - \frac{u}{v} = \frac{v \Delta u - u \Delta v}{(v + \Delta v) v},$$

de la cual

$$\frac{\Delta y}{\Delta x} = \frac{v \frac{\Delta u}{\Delta x} - u \frac{\Delta v}{\Delta x}}{(v + \Delta v) v} \rightarrow \frac{vu' - uv'}{v^2} \quad (\Delta x \rightarrow 0).$$

Aquí hemos utilizado otra vez el hecho de que una función v derivable verifica necesariamente $\Delta v \rightarrow 0$ para $\Delta x \rightarrow 0$. Así:

$$\left(\frac{u}{v}\right)' = \frac{vu' - uv'}{v^2} \quad (15)$$

He aquí algunos ejemplos de aplicación de estas reglas:

$$(2x^3 - 5)' = 2(x^3)' - (5)' = 2 \cdot 3x^2 - 0 = 6x^2;$$

$$(x^2 \sin x)' = x^2 (\sin x)' + (x^2)' \sin x = x^2 \cos x + 2x \sin x;$$

$$\begin{aligned} (\operatorname{tg} x)' &= \left(\frac{\sin x}{\cos x}\right)' = \frac{\cos x (\sin x)' - \sin x (\cos x)'}{\cos^2 x} = \\ &= \frac{\cos x \cdot \cos x - \sin x(-\sin x)}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x. \end{aligned}$$

¹⁴ El último sumando tiende a cero para $\Delta x \rightarrow 0$, puesto que $\Delta v/\Delta x$ tiende a un número finito, igual a la derivada v' , que se supuso existente desde el principio, y $\Delta u \rightarrow 0$, puesto que la función u , por haberse supuesto derivable, es continua.

Recomendamos al lector demuestre por sí mismo la fórmula

$$(\operatorname{ctg} x)' = -\operatorname{cosec}^2 x.$$

Derivada de la función inversa

Consideremos una función $y = f(x)$, que es continua y creciente (decreciente) en el intervalo $[a, b]$. Por función creciente (decreciente)

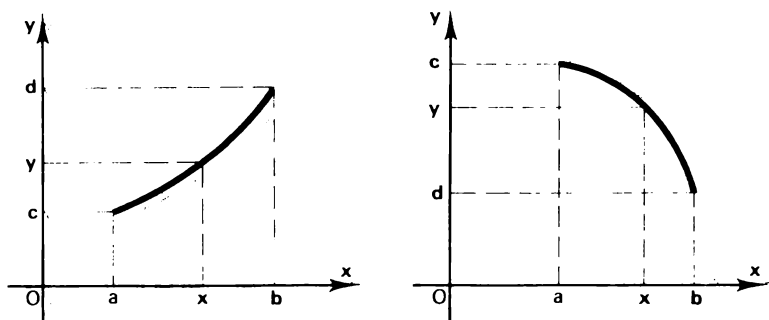


FIG. 14

queremos indicar que, a un incremento positivo de x en el intervalo $[a, b]$, corresponde un incremento positivo (negativo) de y (Fig. 14).

Sea $c = f(a)$ y $d = f(b)$. En la figura 14 es evidente que a cada valor de y del intervalo $[c, d]$ (o $[d, c]$, respectivamente) corresponde exactamente un valor de x del intervalo $[a, b]$ tal que $y = f(x)$. Así, en el intervalo $[c, d]$ (o $[d, c]$) tenemos una función $x = \phi(y)$ perfectamente definida, que se llama *función inversa* de $y = f(x)$. En la figura 14 se ve claramente que la función $\phi(y)$ es continua, hecho que se demuestra en el análisis moderno por métodos estrictamente analíticos. Sean Δx e Δy los incrementos correspondientes a x e y respectivamente. Es evidente que

$$\frac{\Delta y}{\Delta x} = \frac{1}{\Delta x / \Delta y}, \text{ si } \Delta y \neq 0.$$

En el límite esto nos da una sencilla relación entre las derivadas de las funciones directa e inversa

$$y'_x = \frac{1}{x'_y} \cdot \quad (16)$$

Utilicemos esta relación para encontrar la derivada de la función $y = a^x$. La función inversa es $x = \log_a y$, que ya hemos derivado, y así podemos escribir

$$(a^x)'_x = \frac{1}{(\log_a y)'_y} = \frac{1}{\frac{1}{y} (\log_a e)} = y \log_e a = a^x \ln a. \quad (17)$$

En particular $(e^x)' = e^x$.

Otro ejemplo: sea $y = \arcsen x$. La función inversa es $x = \sen y$. Así:

$$(\arcsen x)'_x = \frac{1}{(\sen y)'_y} = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - (\sen y)^2}} = \frac{1}{\sqrt{1 - x^2}}.$$

Tabla de derivadas

Tabulemos las derivadas de las funciones elementales más sencillas (Tabla 2).

TABLA 2

y	y'	y	y'	y	y'
c	0	$\ln x$	$\frac{1}{x}$	$\operatorname{tg} x$	$\sec^2 x$
x^a	ax^{a-1}	$\log_a x$	$\frac{1}{x} \log_a e$	$\arcsen x$	$\frac{1}{\sqrt{1-x^2}}$
e^x	e^x	$\sen x$	$\cos x$	$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$
a^x	$a^x \ln a$	$\cos x$	$-\sen x$	$\operatorname{arctg} x$	$\frac{1}{1+x^2}$

Estas fórmulas ya han sido calculadas y explicadas anteriormente, con la excepción de las dos últimas que el lector puede, si lo desea, deducir fácilmente por sí mismo utilizando la regla de derivación de una función inversa.

Cálculo de la derivada de una función de función

Queda por considerar la última —y más difícil— regla de derivación. El lector que conozca esta regla y tenga a su disposición un conjunto de tablas puede perfectamente considerarse capacitado para derivar cualquier función elemental.

Para aplicar la regla que vamos a dar, es necesario conocer perfectamente cómo está construida la función que deseamos derivar; esto es, qué operaciones debemos realizar sobre la variable independiente x , y en qué orden, para obtener el valor de la variable dependiente y .

Por ejemplo, para calcular la función

$$y = \sin x^2,$$

es necesario en primer lugar elevar x al cuadrado y luego calcular el seno del valor así obtenido, procedimiento que puede describirse del modo siguiente: $y = \sin u$, donde $u = x^2$.

Por otra parte, para calcular la función

$$y = \sin^2 x,$$

es necesario en primer lugar encontrar el seno de x , y luego elevar el valor obtenido al cuadrado, procedimiento que puede escribirse así: $y = u^2$, donde $u = \sin x$.

He aquí algunos ejemplos:

$$1. \quad y = (3x + 4)^3; \quad y = u^3, \quad u = 3x + 4.$$

$$2. \quad y = \sqrt{1 - x^2}; \quad y = u^{1/2}, \quad u = 1 - x^2.$$

$$3. \quad y = e^{kx}; \quad y = e^u, \quad u = kx.$$

En casos más complicados tenemos una cadena de relaciones sencillas con varios eslabones. Por ejemplo,

$$4. \quad y = \cos^3 x^2; \quad y = u^3, \quad u = \cos v, \quad v = x^2.$$

Si y es función de la variable u ,

$$y = f(u), \quad (18)$$

y u es a su vez función de la variable x

$$u = \phi(x), \quad (19)$$

entonces y , al ser función de u , es también función de x , lo que puede notarse como sigue:

$$y = F(x) = f[\phi(x)]. \quad (20)$$

Para considerar casos más complicados podemos formar, por ejemplo, la función

$$y = \Phi(x) = f\{\phi[\psi(x)]\},$$

que es equivalente a las ecuaciones

$$y = f(u), \quad u = \phi(v), \quad v = \psi(x);$$

de este modo podríamos formar cadenas aún más largas.

Veamos ahora cómo calcular la derivada de la función $F(x)$ definida por la ecuación (20), conociendo la derivada de $f(u)$ con respecto a u , y la derivada de $\phi(x)$ con respecto a x .

Demos a x el incremento Δx ; entonces por (19) u se incrementará en Δu y por (18) y se incrementará en Δy . Así podremos escribir

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u} \cdot \frac{\Delta u}{\Delta x}.$$

Hagamos tender Δx a cero. Entonces $\Delta u / \Delta x \rightarrow u'_x$. Además, por la continuidad de u , el incremento $\Delta u \rightarrow 0$, y por tanto $\Delta y / \Delta u \rightarrow y'_u$ (se ha supuesto la existencia de las derivadas y'_u y u'_x).

Hemos demostrado así la importante fórmula para la derivada de la función de función ¹⁵:

$$y'_x = y'_u \cdot u'_x. \quad (21)$$

¹⁵. Para deducir esta fórmula se ha supuesto tácitamente que cuando Δx tiende a cero, Δu no es nunca igual a cero. Pero la fórmula sigue siendo válida aunque esta suposición no se cumpla.

Calculemos, con ayuda de la fórmula (21) y la tabla fundamental de derivadas, las de las funciones que hemos estado considerando:

$$1. \quad y = (3x+4)^3 = u^3; \quad y'_x = (u^3)'_u (3x+4)'_x = 3u^2 \cdot 3 = 9(3x+4)^2$$

$$2. \quad y = \sqrt{1-x^2} = u^{1/2}; \quad y'_x = (u^{1/2})'_u (1-x^2)'_x = \frac{1}{2} u^{-1/2} (-2x) = \\ = -\frac{x}{\sqrt{1-x^2}}$$

$$3. \quad y = e^{kx} = e^u, \quad y'_x = (e^u)'_u (kx)'_x = e^u \cdot k = ke^{kx}.$$

Si $y = f(u)$, $u = \phi(v)$, $v = \psi(x)$, entonces

$$y'_x = y'_u \cdot u'_x = y'_u (u'_v \cdot v'_x) = y'_u \cdot u'_v \cdot v'_x.$$

Se ve claramente cómo debe generalizarse esta fórmula para el caso de que la cadena contenga un número cualquiera (finito) de funciones. Por ejemplo,

$$4. \quad y = \cos^3 x^2; \quad y'_x = (u^3)'_u (\cos v)'_v (x^2)'_x = 3u^2 (-\operatorname{sen} v) 2x = \\ = -6x \cos^2 x^2 \operatorname{sen} x^2.$$

En nuestra explicación de cómo calcular la derivada de una función de función hemos introducido variables intermedias $u, v, \dots$. Sin embargo, en cuanto se adquiere cierta práctica es posible prescindir simplemente de ellas, sin más que retener en la cabeza las funciones que notan.

Las funciones elementales

Para cerrar la presente sección observemos que las funciones cuyas derivadas hemos tabulado (Tabla 2) pueden utilizarse para definir las llamadas funciones elementales. Estas *funciones elementales* se definen como aquellas que pueden obtenerse a partir de las sencillas funciones precedentes por medio de las cuatro operaciones aritméticas y la operación de formar funciones de función, realizadas cada una de estas operaciones un número finito de veces.

Por ejemplo, el polinomio $x^3 - 2x^2 + 3x - 5$ es una función elemental, puesto que se obtiene realizando operaciones aritméticas con un número finito de funciones de la forma x^k . La función

$\ln \sqrt{1 - x^2}$ es también elemental, puesto que se obtiene del polinomio $u = 1 - x^2$, la operación $v = u^{1/2}$, y a continuación la operación $\ln v$.

Las reglas de derivación expuestas son suficientes para obtener la derivada de cualquier función elemental, una vez conocidas las derivadas de las funciones elementales más sencillas.

§7. Máximos y mínimos; estudio de la gráfica de una función

Una de las aplicaciones más sencillas e importantes de la derivada está relacionada con la teoría de los máximos y mínimos. Sea una función $y = f(x)$, definida en un cierto intervalo $a \leq x \leq b$, que no sólo es continua, sino que tiene derivada en todo punto del intervalo. La posibilidad de calcular su derivada permite formar una representación clara de la gráfica de la función. En todo intervalo en el que la derivada sea positiva, la tangente a la gráfica será una recta inclinada hacia la derecha. En un tal intervalo la función será creciente; esto es, a un incremento positivo de x corresponderá un incremento positivo de $f(x)$. Por otra parte, en todo intervalo donde la derivada sea siempre negativa, la función decrecerá y la gráfica disminuirá hacia abajo.

Máximos y mínimos

En la figura 15 hemos dibujado la gráfica de una función $y = f(x)$ definida en el intervalo $[a, b]$. De especial interés son los puntos de esta gráfica cuyas abscisas son x_0, x_1, x_3 .

En el punto x_0 la función $f(x)$ se dice que tiene un máximo relativo, significando con esto que en dicho punto $f(x)$ es mayor que en todos los puntos de un cierto entorno; con más precisión, $f(x_0) \geq f(x)$ para todo x perteneciente a un cierto intervalo alrededor del punto x_0 .

El mínimo relativo se define análogamente.

En la función anterior hay un máximo relativo en los puntos x_0 y x_3 , y un mínimo relativo en el punto x_1 .

En todo punto máximo o mínimo —que sea interior al intervalo $[a, b]$, es decir, que no coincida con ninguno de los puntos extremos a o b —, la derivada debe ser igual a cero.

Esta última afirmación, de suma importancia, se sigue inmediatamente de la definición de derivada como límite del cociente $\Delta y / \Delta x$. En efecto, si nos movemos una pequeña distancia a partir del punto máximo, entonces $\Delta y \leq 0$. Así para Δx positivo el cociente $\Delta y / \Delta x$

es no positivo, y para Δx negativo el cociente $\Delta y/\Delta x$ es no negativo. El límite de este cociente, que existe por hipótesis, no puede ser por tanto positivo ni negativo y sólo queda la posibilidad de que sea cero. Observando el diagrama se ve que esto significa que en los puntos máximo y mínimo (no se suele utilizar la palabra «relativo», aunque se sobreentiende) la tangente a la gráfica es horizontal. En la figura 15 conviene notar que, aunque en los puntos x_2 y x_4 la tangente también

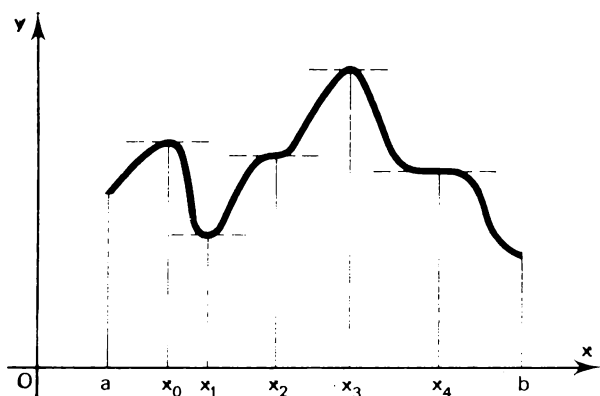


FIG. 15

es horizontal (al igual que en los puntos x_0, x_1, x_3), en estos puntos la función no tiene máximo ni mínimo. En general puede haber más puntos en los que la derivada de la función sea cero (puntos estacionarios) que los que corresponden a máximos o mínimos.

Determinación de los valores máximo y mínimo absolutos de una función

En numerosas cuestiones técnicas es necesario encontrar el punto x en el cual una función dada $f(x)$ toma sus valores mayor (máximo absoluto) y menor (mínimo absoluto) en un intervalo dado.

Si lo que nos interesa es el máximo absoluto, debemos encontrar el punto x_0 del intervalo $[a, b]$ para el cual se verifique la desigualdad $f(x_0) \geq f(x)$ para todos los puntos x del intervalo $[a, b]$.

Surge ahora la cuestión fundamental de si existe siempre un tal punto. Los métodos del análisis moderno permiten demostrar el siguiente teorema de existencia: si la función $f(x)$ es continua en un

intervalo finito, existe al menos un punto del intervalo para el cual la función alcanza su valor máximo absoluto (mínimo absoluto) en el intervalo $[a, b]$.

De lo dicho se deduce que estos puntos máximo o mínimo deben buscarse entre los puntos «estacionarios». Este hecho es la base del siguiente método general de obtención de los máximos y mínimos.

En primer lugar se calcula la derivada de $f(x)$, y luego resolvemos la ecuación que se obtiene igualando a cero:

$$f'(x) = 0.$$

Si $x_1, x_2, \dots, x_n$ son las raíces de esta ecuación, comparamos entonces los números $f(x_1), f(x_2), \dots, f(x_n)$ entre sí. Naturalmente, es necesario tener en cuenta que el máximo o mínimo de la función puede no encontrarse dentro del intervalo, sino en el extremo (como ocurre con el mínimo de la figura 15) o en un punto donde la función no tiene derivada (como en la figura 12). Así, pues, a los puntos $x_1, x_2, \dots, x_n$ es preciso añadir los extremos a y b del intervalo y también aquellos puntos —si existen— en los que no hay derivada. Sólo queda comparar los valores de la función en todos estos puntos y elegir entre ellos el mayor o el menor.

Con respecto al teorema de existencia establecido, es importante añadir que este teorema deja, en general, de verificarse en el caso de que la función $f(x)$ sea continua sólo en el intervalo abierto (a, b) ; esto es, en el conjunto de puntos x que satisfacen la desigualdad $a < x < b$. Dejamos al lector que compruebe que la función $1/x$ no tiene ni máximo ni mínimo en el intervalo $(0, 1)$.

Analicemos algunos ejemplos.

De una plancha cuadrada de estaño de lado a se pide construir una caja rectangular abierta de volumen máximo. Si de las esquinas del cuadrado original separamos cuadrados de lado x (ver §2, ejemplo 2) obtenemos una caja de volumen

$$V = x(a - 2x)^2.$$

El problema consiste entonces en encontrar el valor de x para el cual la función $V(x)$ alcanza su máximo absoluto en el intervalo $0 \leq x \leq a/2$. De acuerdo con la regla establecida, calculamos la derivada y la igualamos a cero:

$$V'(x) = (a - 2x)^2 - 4x(a - 2x) = 0.$$

Resolviendo esta ecuación, encontramos las dos raíces:

$$x_1 = \frac{a}{2}, \quad x_2 = \frac{a}{6}.$$

A éstas añadimos el extremo inferior del intervalo (el extremo superior es x_1) y comparamos los valores de la función en estos puntos

$$V(0) = 0; \quad V\left(\frac{a}{6}\right) = \frac{2}{27} a^3; \quad V\left(\frac{a}{2}\right) = 0.$$

Así la caja tendrá un volumen máximo igual a $\frac{2}{27} a^3$, para la altura $x = a/6$.

Como segundo ejemplo examinemos el problema de la farola en la pista de patinaje (ver §2, ejemplo 3). ¿A qué altura h deberá situarse la lámpara para que el borde de la pista reciba la máxima iluminación?

Teniendo en cuenta la fórmula (3) del §2, el problema se reduce a determinar el valor de h para que $T = (A \operatorname{sen} \alpha)/(h^2 + r^2)$ alcance su máximo absoluto. En lugar de h es más conveniente obtener el ángulo α (Fig. 3. Cap. 2). Tenemos

$$h = r \operatorname{tg} \alpha,$$

por lo que

$$T = \frac{A}{r^2} \frac{\operatorname{sen} \alpha}{1 + \operatorname{tg}^2 \alpha} = \frac{A}{r^2} \operatorname{sen} \alpha \cos^2 \alpha.$$

Se pide entonces obtener el máximo absoluto de la función $T(\alpha)$ entre aquellos valores de α que satisfacen la desigualdad $0 < \alpha < \pi/2$. Para ello obtenemos la derivada y la igualamos a cero:

$$T'(\alpha) = \frac{A}{r^2} (\cos^3 \alpha - 2 \operatorname{sen}^2 \alpha \cos \alpha) = 0.$$

Esta ecuación se separa en dos:

$$\cos \alpha = 0; \quad \cos^2 \alpha - 2 \operatorname{sen}^2 \alpha = 0.$$

La primera ecuación tiene la raíz $\alpha = \pi/2$, que coincide con el extremo del intervalo $(0, \pi/2)$. La segunda puede ponerse en la forma

$$\operatorname{tg}^2 \alpha = \frac{1}{2}.$$

Pero dado que $0 < \alpha < \pi/2$, tenemos el resultado $\alpha \approx 35^\circ 15'$. Este es el valor para el cual $T(\alpha)$ alcanza su máximo absoluto (en los extremos del intervalo, $T = 0$). La altura pedida h es igual a:

$$h = r \operatorname{tg} \alpha = \frac{r}{\sqrt{2}} \approx 0,7 r.$$

Para que la iluminación del borde de la pista sea óptima, la lámpara debe situarse a una altura aproximadamente igual a 0,7 veces el radio.

Pero supongamos ahora que los medios a nuestra disposición

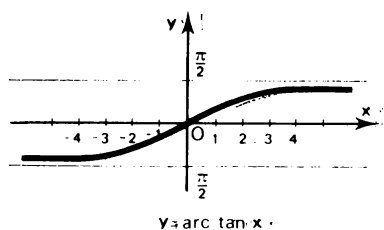


FIG. 16 a

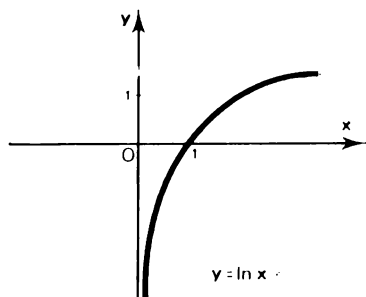


FIG. 16 b

no nos permiten elevar la lámpara hasta una altura superior a un cierto H . Entonces el ángulo α no puede variar de 0 a $\pi/2$, sino sólo dentro de los límites $0 < \alpha \leq \operatorname{arc} \operatorname{tg} (H/r)$. Por ejemplo, sea r igual a 12 metros y H igual a 9 metros. En este caso es posible, en efecto, elevar la lámpara a la altura $h = r/\sqrt{2}$, que supone algo más de 8 metros, y esto es lo que se haría. Pero si H es menor quince metros (por ejemplo, si únicamente tenemos a nuestra disposición un poste de 6 metros de longitud), entonces resulta que la derivada de la función $T(\alpha)$ en el intervalo $[0, \operatorname{arc} \operatorname{tg} (H/r)]$ no es nunca igual a cero. En este caso el máximo se alcanza en el extremo del intervalo, y la lámpara deberá ser elevada a la mayor altura posible, es decir, $H = 6$ metros.

Hasta ahora sólo hemos considerado funciones sobre un intervalo finito. Si el intervalo es de longitud infinita, puede ocurrir que la función continua no alcance nunca valores máximo o mínimo, pues

puede, por ejemplo, crecer o decrecer indefinidamente cuando x tiende a infinito.

Así, las funciones $y = kx + b$ (ver Fig. 5, Cap. 2), $y = \arctg x$ (Fig. 16a), $y = \ln x$ (Fig. 16b), nunca alcanzan su valor máximo o mínimo. La función $y = e^{-x^2}$ (Fig. 16c) alcanza su valor máximo en el punto $x = 0$, pero no tiene mínimo. La función $y = x/(1 + x^2)$ (Fig. 16d) alcanza su valor mínimo en el punto $x = -1$, y su máximo, en el punto $x = 1$.

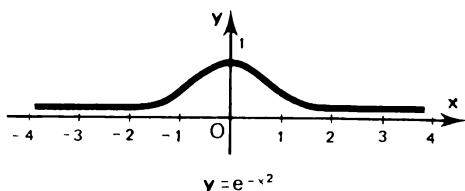


FIG. 16 c

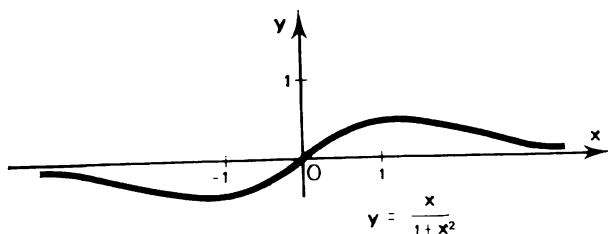


FIG. 16 d

El caso de un intervalo de longitud infinita puede ser reducido al caso ordinario sin más que considerar, en lugar de $f(a)$ y $f(b)$, los límites

$$A = \lim_{x \rightarrow -\infty} f(x), \quad B = \lim_{x \rightarrow +\infty} f(x).$$

Derivadas de órdenes superiores

Acabamos de ver que para efectuar un estudio más a fondo de la gráfica de una función, se debe estudiar su derivada $f'(x)$. Esta derivada es una función de x , por lo que puede obtenerse a su vez su derivada.

Su derivada se llama derivada segunda y se nota por

$$[y']' = y'' \quad \text{o} \quad [f'(x)]' = f''(x).$$

Análogamente, podemos calcular la derivada tercera

$$[y'']' = y''' \quad \text{ó} \quad [f''(x)]' = f'''(x)$$

y en general la derivada n -ésima o, como también se dice, la derivada de orden n :

$$y^{(n)} = f^{(n)}(x).$$

Naturalmente, debe tenerse en cuenta que para un cierto valor de x (e incluso para todos los valores de x) la función puede no tener derivadas de algún orden; por ejemplo, la k -ésima; puede suceder que exista $f^{(k)}(x)$, pero no $f^{(k+1)}(x)$. Las derivadas de órdenes superiores aparecerán más tarde en el §9 en conexión con la fórmula de Taylor. Por el momento nos limitaremos a la derivada segunda.

Significado de la derivada segunda; convexidad y concavidad

La derivada segunda tiene una interpretación muy sencilla en mecánica. Sea $s = f(t)$ la ecuación de movimiento a lo largo de una línea recta; entonces s' es la velocidad y s'' es el «ritmo de cambio de la velocidad», o sencillamente la «aceleración» del punto en el

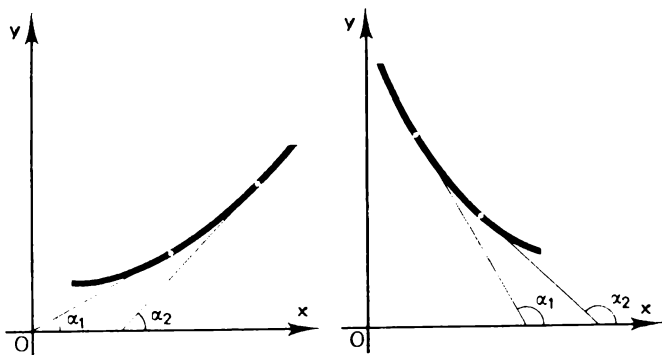


FIG. 17

instante t . Por ejemplo, para un cuerpo que cae bajo la acción de la gravedad tenemos:

$$s = \frac{gt^2}{2} + v_0 t + s_0,$$

$$s' = gt + v_0,$$

$$s'' = g,$$

esto es, la aceleración de los cuerpos en caída libre es constante.

La derivada segunda tiene también un significado geométrico sencillo. Al igual que el signo de la derivada primera determina si la función es creciente o decreciente, así el signo de la derivada segunda determina el lado hacia el cual se curva la gráfica de la función.

Supongamos, por ejemplo, que sobre un intervalo dado la derivada segunda es siempre positiva. Entonces la derivada primera es creciente, y por tanto $f'(x) = \operatorname{tg} \alpha$ aumenta, lo mismo que el ángulo α de inclinación de la recta tangente (Fig. 17). Así, si nos desplazamos a lo largo de la gráfica, ésta se curva constantemente hacia un mismo lado, hacia arriba, y se dice que es «convexa hacia abajo».

Por otro lado, en una parte de la curva en que la derivada segunda sea constantemente negativa (Fig. 18), la gráfica de la función es «convexa hacia arriba»¹⁶.

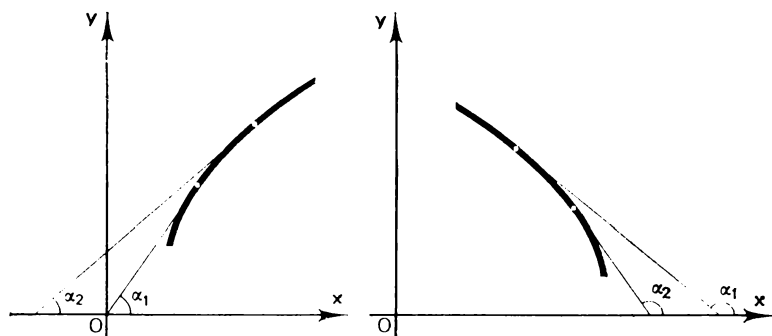


FIG. 18

Criterios de máximo y mínimo; estudio de la gráfica de una curva

Si en todo el intervalo de variación x la curva es convexa hacia arriba y si en un cierto punto x_0 de este intervalo la derivada es igual a cero, entonces en este punto la función necesariamente tiene un máximo (y un mínimo en el caso de que la convexidad sea hacia abajo). Esta sencilla consideración permite decidir a menudo —una

¹⁶ Definida estrictamente, la «convexidad hacia arriba» es aquella propiedad de la curva que consiste en encontrarse por encima (con más precisión «no por debajo») de la cuerda que une dos cualesquiera de sus puntos; análogamente la «convexidad hacia abajo» (que también se llama sencillamente «concavidad»), significa que la curva no se encuentra por encima de cualquiera de sus cuerdas.

vez encontrado un punto en el que la derivada sea cero— si en este punto la función tiene un máximo o un mínimo relativo¹⁷.

Ejemplo 1

Estudiamos la forma de la gráfica de la función

$$f(x) = \frac{x^3}{3} - \frac{5x^2}{2} + 6x - 2.$$

Se calcula su primera derivada y se iguala a cero:

$$f'(x) = x^2 - 5x + 6 = 0.$$

Las raíces de la ecuación obtenida de este modo son $x_1 = 2$, $x_2 = 3$. Los valores correspondientes de la función son

$$f(2) = 2\frac{2}{3}, \quad f(3) = 2\frac{1}{2}.$$

Marcamos entonces estos dos puntos sobre el diagrama. Junto con éstos podemos señalar también el punto de coordenadas $x = 0$ e $y = f(0) = -2$, donde la gráfica corta al eje y . La derivada segunda es $f''(x) = 2x - 5$. Esta se anula para $x = 5/2$, por lo que

$$f''(x) > 0 \text{ para } x > \frac{5}{2},$$

$$f''(x) < 0 \text{ para } x < \frac{5}{2}.$$

El punto

$$x = \frac{5}{2}, \quad y = f\left(\frac{5}{2}\right) = 2\frac{7}{12}$$

es un *punto de inflexión* de la gráfica. A la izquierda de este punto la curva es convexa hacia arriba, y a la derecha es convexa hacia abajo.

Ahora resulta evidente que en el punto $x = 2$ hay un máximo y en el punto $x = 3$ un mínimo de la función.

¹⁷ En casos más complicados, cuando la derivada segunda cambia de signo, el problema de determinar el carácter de un punto estacionario se resuelve por medio de la fórmula de Taylor (§9).

De estos resultados concluimos que la gráfica de la función $y = f(x)$ es la esquematizada en la figura 19. A la derecha del punto $(0, -2)$ la curva crece con la x , es convexa hacia arriba y alcanza su máximo en el punto $(2, 2\frac{2}{3})$, después del cual comienza a decrecer. En el punto $(2\frac{1}{2}, 2\frac{7}{12})$, donde $f''(x) = 0$, la convexidad cambia a concavidad. En el punto $(3, 2\frac{1}{2})$ la función tiene un mínimo, y a partir de allí crece indefinidamente. Este último resultado se deduce del hecho de que el primer término de la función —el que contiene la mayor potencia (tercera) de x — tiende a infinito más rápidamente que los términos segundo y tercero. Por la misma razón la gráfica tiende a $-\infty$ cuando x toma valores negativos grandes (en valor absoluto).

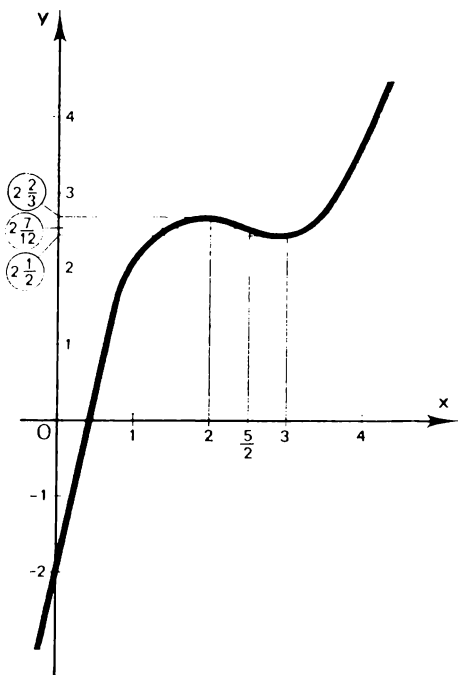


FIG. 19

Ejemplo 2

Demostremos la desigualdad $e^x \geq 1 + x$ para todo x . Para ello consideremos la función $f(x) = e^x - x - 1$. Su primera derivada es $f'(x) = e^x - 1$, que se anula únicamente para $x = 0$. La derivada segunda es $f''(x) = e^x > 0$

para todo x . Por tanto la gráfica de la función $f(x)$ es convexa hacia abajo. El número $f(0) = 0$ es un mínimo de la función, y para todo x se verifica $e^x - x - 1 \geq 0$.

El estudio de las gráficas tiene interés por muchas razones. A menudo muestran muy claramente, por ejemplo, el número de raíces reales de una ecuación dada. Así, para demostrar que la ecuación

$$xe^x = 2$$

tiene una única raíz real, podemos estudiar la gráfica de las funciones $y = e^x$, $y = 2/x$ (dibujadas en la figura 20). Es fácil ver que estas gráficas se cortan en un solo punto, por lo que la ecuación $e^x = 2/x$ tiene exactamente una raíz.

Los métodos del análisis se aplican extensamente al cálculo aproximado de las raíces de una ecuación. Para esta aplicación véase el capítulo 4, §5.

§8. Incremento y diferencial de una función

La diferencial de una función

Consideremos una función derivable $y = f(x)$. El incremento de esta función

$$\Delta y = f(x + \Delta x) - f(x)$$

correspondiente al incremento Δx , tiene la propiedad de que el cociente $\Delta y / \Delta x$, cuando $\Delta x \rightarrow 0$, tiende a un límite finito, igual a la derivada

$$\frac{\Delta y}{\Delta x} \rightarrow f'(x).$$

Esto puede escribirse en forma de igualdad

$$\frac{\Delta y}{\Delta x} = f'(x) + \alpha,$$

donde el valor de α depende de Δx de modo tal que cuando $\Delta x \rightarrow 0$, α también tiende a cero. Así el incremento de una función puede representarse en la forma

$$\Delta y = f'(x) \Delta x + \alpha \Delta x,$$

donde $\alpha \rightarrow 0$, para $\Delta x \rightarrow 0$.

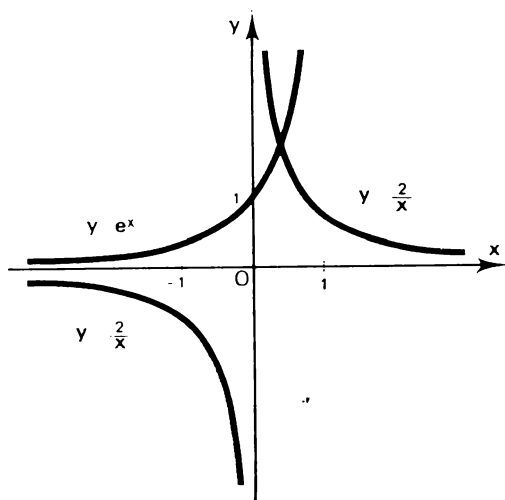


FIG. 20

El primer sumando del segundo miembro de esta igualdad depende de Δx de un modo muy sencillo: es proporcional a Δx . Se llama la *diferencial* de la función, en el punto x , correspondiente al incremento Δx , y se nota por

$$dy = f'(x) \Delta x.$$

El segundo sumando tiene la propiedad característica de que cuando $\Delta x \rightarrow 0$, tiende a cero más rápidamente que Δx , como resultado de la presencia del factor α . Se dice así que es un infinitésimo de orden superior a Δx y, en el caso $f'(x) \neq 0$, es también de orden superior al del primer sumando. Esto significa que, para Δx suficientemente pequeño, el segundo sumando se puede hacer tan pequeño como se quiera y que su cociente, al dividirlo por Δx , también tiene la misma propiedad.

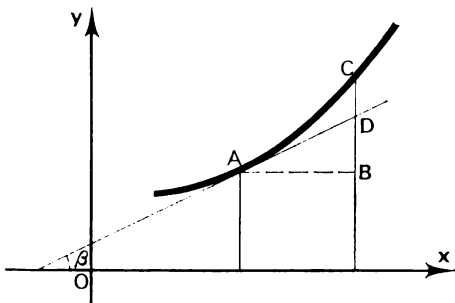


FIG. 21

La descomposición de Δy en dos sumandos, de los cuales el primero (la parte principal) depende linealmente de Δx y el segundo es despreciable para Δx pequeño, está ilustrada en la figura 21. El segmento $BC = \Delta y$, donde $BC = BD + DC$; $BD = \operatorname{tg} \beta \cdot \Delta x = f'(x) \cdot \Delta x = dy$ y DC es un infinitésimo de orden superior a Δx .

En la práctica, la diferencial se emplea a menudo como un valor aproximado del incremento de la función. Por ejemplo, supongamos que se plantea el problema de determinar el volumen de las paredes de una caja cúbica cerrada cuyas dimensiones interiores son $10 \times 10 \times 10$ cm y cuyo espesor es 0,05 cm. Si no se necesita gran exactitud, podemos razonar como sigue. El volumen limitado por las paredes interior y exterior de la caja representa el incremento Δy de la función $y = x^3$ para $x = 10$ y $\Delta x = 0,1$. Así encontramos aproximadamente

$$\Delta y \approx dy = (x^3)' \Delta x = 3x^2 \Delta x = 3 \cdot 10^2 \cdot (0,1) = 30 \text{ cm}^3.$$

Con el fin de alcanzar cierta simetría en la notación se acostumbra notar el incremento de la variable independiente por dx y llamarlo también diferencial. Con esta notación la diferencial de la función puede escribirse así:

$$dy = f'(x) dx.$$

Entonces la derivada es el cociente $f'(x) = dy/dx$ entre la diferencial de la función y la diferencial de la variable independiente.

La diferencial de una función surgió históricamente del concepto de «indivisible». Este concepto, que desde un punto de vista moderno nunca estuvo muy claramente definido, era en su tiempo —en el siglo XVIII— fundamental en el análisis matemático. Las ideas referentes a él sufrieron cambios esenciales en el transcurso de varios siglos. Los indivisibles, y más tarde la diferencial de una función, se representaban como verdaderos infinitésimos, como algo de magnitud constante extremadamente pequeña, que sin embargo no era cero. La definición dada en esta sección es la aceptada en el análisis moderno. De acuerdo con esta definición, la diferencial es una magnitud finita para cada incremento Δx , y al mismo tiempo proporcional a Δx . La otra propiedad fundamental de la diferencial, el carácter de su diferencia respecto a Δy , sólo puede reconocerse «en movimiento», por así decirlo: si consideramos un incremento Δx que se aproxima a cero (que sea un infinitésimo), entonces la diferencia entre dy e Δy será tan pequeña como se desee incluso comparada con Δx .

Esta sustitución de los incrementos pequeños de la función por la diferencial forma la base de la mayoría de las aplicaciones del análisis infinitesimal al estudio de la naturaleza. El lector verá esto de un modo particularmente claro en el caso de las ecuaciones diferenciales, de las que se tratará en los capítulos 5 y 6.

Así, para determinar la función que representa un proceso físico dado, se intenta en primer lugar obtener una ecuación que relacione de algún modo esta función con sus derivadas de distintos órdenes. El método para obtener una tal ecuación, llamada ecuación diferencial, consiste a menudo en reemplazar los incrementos de las funciones buscadas por sus diferenciales correspondientes.

Como ejemplo resolvamos el siguiente problema. En un sistema de coordenadas rectangulares $Oxyz$, consideremos la superficie obtenida por rotación de la parábola cuya ecuación (en el plano Oyz) es $z = y^2$. La superficie obtenida se llama paraboloide de revolución (Fig. 22). Notemos por v el volumen del cuerpo limitado

por el paraboloide y el plano paralelo al plano Oxy a distancia z de él. Es evidente que v es función de z ($z > 0$).

Para determinar la función v , tratemos de encontrar su diferencial dv . El incremento Δv de la función v en el punto z es igual al volumen limitado por el paraboloide y por dos planos paralelos al Oxy a distancias z y $z + \Delta z$ de él.

Se comprueba fácilmente que el valor de Δv es mayor que el volumen del cilindro circular de radio $\sqrt{z}$ y altura Δz , pero menor que el del cilindro circular de radio $\sqrt{z + \Delta z}$ y altura Δz .

Así

$$\pi z \Delta z < \Delta v < \pi(z + \Delta z) \Delta z$$

y por tanto

$$\Delta v = \pi(z + \theta \Delta z) \Delta z = \pi z \Delta z + \pi \theta \Delta z^2,$$

donde θ es un número que depende de Δz y que satisface la desigualdad $0 < \theta < 1$.

Hemos logrado así representar el incremento Δv en forma de una suma, cuyo primer sumando es proporcional a Δz , mientras el segundo es un infinitésimo de orden superior a Δz (cuando $\Delta z \rightarrow 0$). Se deduce por tanto que el primer sumando es la diferencial de la función v

$$dv = \pi z \Delta z,$$

o bien

$$dv = \pi z dz,$$

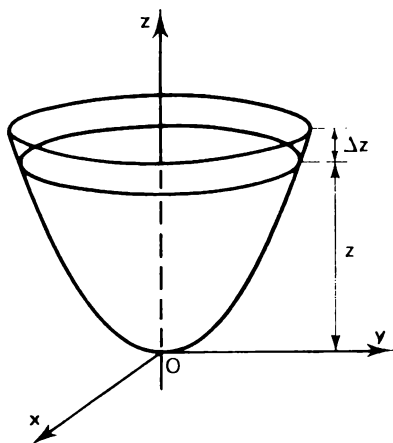


FIG. 22

puesto que $\Delta z = dz$ para la variable independiente z .

La ecuación obtenida relaciona las diferenciales dv y dz (de las variables v y z) entre sí y se dice que es una ecuación diferencial.

Teniendo en cuenta que

$$\frac{dv}{dz} = v',$$

donde v' es la derivada de v con respecto a la variable z , la ecuación diferencial puede también escribirse en la forma

$$v' = \pi z.$$

Para resolver esta sencilla ecuación diferencial debemos encontrar una función de z cuya derivada sea igual a πz . Problemas de este tipo se estudiarán de modo general en las secciones §10 y §11, pero por el momento pedimos al lector que compruebe que una solución de la ecuación viene dada por $v = \pi z^2/2 + C$, donde C puede ser cualquier número¹⁸. En este caso el volumen del cuerpo es evidentemente cero para $z = 0$ (ver Fig. 22), por lo que $C = 0$. Así, la función viene dada por $v = \pi z^2/2$.

El teorema del valor medio y ejemplos de aplicación

La diferencial expresa el valor aproximado del incremento de la función en términos del incremento de la variable independiente y de la derivada en el punto inicial. Así, para el incremento desde $x = a$ a $x = b$, tenemos

$$f(b) - f(a) \approx f'(a)(b - a).$$

Se puede obtener una verdadera igualdad si se reemplaza la derivada $f'(a)$ en el punto inicial por la derivada en un punto intermedio, convenientemente elegido, del intervalo (a, b) . Precizando: *Si $y = f(x)$ es una función diferenciable en el intervalo $a \leq x \leq b$, existe un punto ξ , contenido estrictamente en este intervalo, tal que se verifica la igualdad*

$$f(b) - f(a) = f'(\xi)(b - a). \quad (22)$$

La interpretación geométrica de este «teorema del valor medio» (también llamado fórmula de Lagrange o fórmula de los incrementos finitos) es extraordinariamente sencilla. Sean A y B los puntos de la gráfica de la función $f(x)$ que corresponden a $x = a$ y $x = b$, y unamos A y B con la cuerda AB (Fig. 23). Movamos ahora la recta AB hacia arriba o hacia abajo manteniéndola constantemente paralela a sí misma. En el punto en el que esta recta corte a la gráfica por última vez, será tangente a ella en un cierto punto C . En este punto (sea $x = \xi$ la abscisa correspondiente), la recta tangente tendrá el

¹⁸ Esta fórmula da todas las soluciones.

mismo ángulo de inclinación α que la cuerda AB . Pero para la cuerda tenemos

$$\operatorname{tg} \alpha = \frac{f(b) - f(a)}{b - a}.$$

Por otra parte, en el punto C es:

$$\operatorname{tg} \alpha = f'(\xi).$$

Esta ecuación

$$\frac{f(b) - f(a)}{b - a} = f'(\xi)$$

es justamente el teorema del valor medio¹⁹.

La fórmula (22) tiene de particular el que el punto ξ que aparece en ella es desconocido; sólo sabemos que se encuentra «dentro del

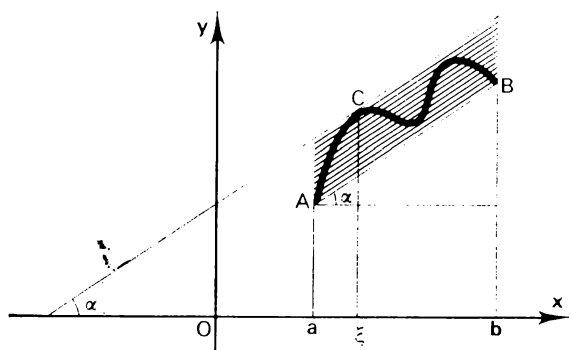


FIG. 23

intervalo (a, b) ». Pero, a pesar de esta indeterminación, la fórmula tiene gran importancia teórica y se utiliza en la demostración de muchos teoremas del análisis. La importancia práctica de esta fórmula es también muy grande, puesto que permite estimar el creci-

¹⁹ Naturalmente, estos razonamientos sólo proporcionan una interpretación geométrica del teorema y de ningún modo constituyen una demostración rigurosa.

miento de una función cuando conocemos los límites entre los cuales puede variar su derivada. Por ejemplo,

$$|\sin b - \sin a| = |\cos \xi| (b - a) \leq b - a.$$

En esta fórmula a , b y ξ son ángulos, expresados en radianes; ξ es un número comprendido entre a y b , pero desconocido, aunque sabemos que $|\cos \xi| \leq 1$.

De la fórmula 22 se deduce que una función cuya derivada es en todo punto igual a cero debe ser constante; en ninguna parte del intervalo puede recibir un incremento no nulo. Análogamente el lector demostrará fácilmente que una función cuya derivada es positiva en todo punto debe ser siempre creciente; y que, si su derivada es negativa, la función debe decrecer. Damos ahora, sin demostración, una de las muchas generalizaciones del teorema del valor medio.

*Para cualesquiera funciones $\phi(x)$ y $\psi(x)$ diferenciables en el intervalo $[a, b]$, con la única restricción de que $\psi'(x) \neq 0$ para todo x de (a, b) , se verifica la ecuación*²⁰

$$\frac{\phi(b) - \phi(a)}{\psi(b) - \psi(a)} = \frac{\phi'(\xi)}{\psi'(\xi)}, \quad (23)$$

donde ξ es cierto punto del intervalo (a, b) ²¹.

De este teorema podemos obtener un método general para el cálculo de límites de expresiones como la siguiente:

$$\lim_{x \rightarrow 0} \frac{\phi(x)}{\psi(x)}, \quad (24)$$

siendo $\phi(0) = \psi(0) = 0$. Por la fórmula (23), tenemos

$$\frac{\phi(x)}{\psi(x)} = \frac{\phi(x) - \phi(0)}{\psi(x) - \psi(0)} = \frac{\phi'(\xi)}{\psi'(\xi)},$$

²⁰ La fórmula (23) puede deducirse por una sencilla aplicación del teorema del valor medio a la función

$$f(x) = \phi(x) - \frac{\phi(b) - \phi(a)}{\psi(b) - \psi(a)} \psi(x)$$

²¹ Por los símbolos $[a, b]$ y (a, b) notamos los conjuntos de valores de x que verifican las desigualdades $a \leq x \leq b$ y $a < x < b$, respectivamente.

donde ξ está comprendido entre 0 y x , y por tanto $\xi \rightarrow 0$ simultáneamente con x . Esto permite calcular el límite

$$\lim_{x \rightarrow 0} \frac{\phi'(x)}{\psi'(x)},$$

en lugar del límite (24), que es en muchos casos más difícil de obtener ²².

Ejemplo

Calculemos $\lim_{x \rightarrow 0} \frac{x - \operatorname{sen} x}{x^3}$. Aplicando tres veces la regla anterior,

tenemos sucesivamente

$$\lim_{x \rightarrow 0} \frac{x - \operatorname{sen} x}{x^3} = \lim_{x \rightarrow 0} \frac{1 - \cos x}{3x^2} = \lim_{x \rightarrow 0} \frac{\operatorname{sen} x}{6x} = \lim_{x \rightarrow 0} \frac{\cos x}{6} = \frac{1}{6}.$$

§9. Fórmula de Taylor

La función

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n,$$

en la que los coeficientes a_k son constantes, se llama polinomio de grado n . En particular, $y = ax + b$ es un polinomio de primer grado e $y = ax^2 + bx + c$ es un polinomio de segundo grado. Los polinomios pueden considerarse las funciones más sencillas de todas. Para calcular su valor para una x dada, necesitamos emplear únicamente las operaciones de adición, substracción y multiplicación; ni siquiera la división es necesaria. Los polinomios son funciones continuas para todo x y tienen derivadas de cualquier orden. Además, la derivada de un polinomio es también un polinomio de grado inferior en una unidad, y las derivadas de orden $n + 1$ y superiores de un polinomio de grado n son nulas.

²² La misma regla es válida para encontrar el límite de una expresión fraccionaria en la que tanto el numerador como el denominador tienden a infinito. Este método, que es muy conveniente para encontrar tales límites (o, como se suele decir, para eliminar la indeterminación), se utilizará, por ejemplo, en el §3 del capítulo 12.

Si a los polinomios añadimos las funciones de la forma

$$y = \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m}$$

(para cuyo cálculo necesitamos también de la división), las funciones $\sqrt{x}$ y $\sqrt[3]{x}$, y, finalmente, las combinaciones aritméticas de los tipos anteriores, obtenemos esencialmente las funciones cuyos valores pueden calcularse por métodos aprendidos en el bachillerato.

A este nivel se tienen nociones de algunas otras funciones tales como

$$\sqrt[5]{x}, \log x, \operatorname{sen} x, \operatorname{arc} \operatorname{tg} x, \dots$$

pero, aunque se estudian sus propiedades más importantes, no se da una respuesta en la matemática elemental a la pregunta: ¿Cómo calcularlas? ¿Qué clase de operaciones, por ejemplo, es necesario realizar sobre la x para obtener $\log x$ o $\operatorname{sen} x$? La respuesta a esta pregunta la proporcionan los métodos desarrollados por el análisis. Examinemos uno de estos métodos.

Fórmula de Taylor

Sea $f(x)$ una función definida en un intervalo que contiene al punto a , con derivada de todos los órdenes.

El polinomio de primer grado

$$p_1(x) = f(a) + f'(a)(x - a)$$

tiene el mismo valor que $f(x)$ en el punto $x = a$ y también, como se comprueba fácilmente, la misma derivada que $f(x)$ en este punto. Su gráfica es una recta tangente a la gráfica de $f(x)$ en el punto a . Es posible elegir un polinomio de segundo grado,

$$p_2(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2,$$

tal que en el punto $x = a$ tenga el mismo valor que $f(x)$ y valores también iguales para su primera y segunda derivadas. Su gráfica en el punto a se acercará a la de $f(x)$ más que la anterior. Es natural esperar que si construimos un polinomio que en $x = a$ tenga las mismas n primeras derivadas que $f(x)$ en el mismo punto, este poli-

nomio se aproximará todavía más a $f(x)$ en los puntos x próximos a a . Así obtenemos la siguiente igualdad aproximada, que es la fórmula de Taylor:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n. \quad (25)$$

El segundo miembro de esta fórmula es un polinomio de grado n en $(x-a)$. Para cada x puede calcularse el valor de este polinomio si se conocen los valores de $f(a), f'(a), \dots, f^{(n)}(a)$.

Para funciones que tienen derivada $(n+1)$ -ésima, el segundo miembro de esta fórmula, como se demuestra fácilmente, difiere del primero en una pequeña cantidad que tiende a cero más rápidamente que $(x-a)^n$. Además, es el único polinomio de grado n que difiere de $f(x)$, para x próximo a a , en un valor que tiende a cero (cuando $x \rightarrow a$) más rápidamente que $(x-a)^n$. Si $f(x)$ es un polinomio algebraico de grado n , entonces la igualdad aproximada (25) es una verdadera igualdad.

Por último, y esto es particularmente importante, podemos dar una sencilla expresión para la diferencia entre el segundo miembro de la fórmula (25) y el verdadero valor de $f(x)$. Para que sea exacta la igualdad aproximada (25), debemos añadir al segundo miembro un término más, llamado «resto»:

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}. \quad (26)$$

Este último término suplementario ²³

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

tiene la peculiaridad de que la derivada que aparece en él debe calcularse en cada caso, no en el punto a , sino en un punto ξ convenientemente elegido, desconocido, pero interior al intervalo de extremos a y x .

La demostración de la igualdad (26) es bastante engorrosa, aun-

²³ Esta es sólo una de las posibles formas del resto $R_{n+1}(x)$.

que sencilla en esencia. Daremos aquí una versión algo artificiosa de la demostración, que tiene la ventaja de ser concisa.

Para encontrar la diferencia entre el primer miembro y el segundo en la fórmula aproximada (25), consideremos el cociente cuyo numerador es la diferencia entre los dos miembros de la igualdad (25) y cuyo denominador es la expresión $-(x-a)^{n+1}$

$$\frac{f(x) - \left[f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!} (x-a)^n \right]}{-(x-a)^{n+1}}. \quad (27)$$

Introduzcamos también la función

$$\phi(u) = f(u) + f'(u)(x-u) + \cdots + \frac{f^{(n)}(u)}{n!} (x-u)^n$$

de la variable u , considerando x fija (constante). El numerador de (27) representará entonces el aumento de esta función cuando pasamos de $u = a$ a $u = x$, y el denominador será el aumento, en ese mismo intervalo, de la función

$$\psi(u) = (x-u)^{n+1}.$$

Utilicemos ahora el teorema del valor medio generalizado que enunciamos anteriormente:

$$\frac{\phi(x) - \phi(a)}{\psi(x) - \psi(a)} = \frac{\phi'(\xi)}{\psi'(\xi)}.$$

Derivando las funciones $\phi(u)$ y $\psi(u)$ con respecto a u (debe recordarse que el valor de x se considera fijo), encontramos que

$$\frac{\phi'(\xi)}{\psi'(\xi)} = -\frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

La igualdad entre esta última expresión y la cantidad primitiva (27) da la fórmula de Taylor en la forma (26).

En la forma (26), la fórmula de Taylor no sólo proporciona un método de calcular aproximadamente $f(x)$, sino que también permite estimar el error. Consideremos el sencillo ejemplo

$$y = \text{sen } x.$$

Los valores de la función $\text{sen } x$ y de sus derivadas de cualquier orden para $x = 0$ son conocidos. Utilicemos estos valores para escribir la fórmula de Taylor para $\text{sen } x$, tomando $a = 0$ y limitándonos al caso $n = 4$. Encontramos sucesivamente:

$$\begin{aligned} f(x) &= \text{sen } x, & f'(x) &= \cos x, & f''(x) &= -\text{sen } x, \\ f'''(x) &= -\cos x, & f^{IV}(x) &= \text{sen } x, & f^V(x) &= \cos x; \\ f(0) &= 0, & f'(0) &= 1, & f''(0) &= 0, \\ f'''(0) &= -1, & f^{IV}(0) &= 0, & f^V(\xi) &= \cos \xi. \end{aligned}$$

Por tanto,

$$\text{sen } x = x - \frac{x^3}{6} + R_5, \text{ donde } R_5 = \frac{x^5}{120} \cos \xi.$$

Aunque se desconoce el valor exacto de R_5 , puede ser fácilmente estimado por el hecho de que $|\cos \xi| \leq 1$. Para todos los valores de x entre cero y $\pi/4$ tenemos

$$|R_5| = \left| \frac{x^5}{120} \cos \xi \right| < \frac{1}{120} \left(\frac{\pi}{4} \right)^5 < \frac{1}{400}.$$

Por tanto, en el intervalo $[0, \pi/4]$ la función $\text{sen } x$ puede calcularse con una aproximación de hasta $1/400$ mediante el polinomio de tercer grado

$$\text{sen } x = x - \frac{1}{6} x^3.$$

Si tomásemos más términos en el desarrollo de Taylor de $\text{sen } x$, obtendríamos un polinomio de grado superior que se aproximaría más todavía a $\text{sen } x$.

Las tablas para las funciones trigonométricas y otras muchas se calculan por métodos semejantes.

Las leyes naturales pueden expresarse, por regla general, con buena aproximación por funciones derivables un número arbitrario de veces, y por ello pueden ser aproximadas por polinomios cuyo grado viene determinado por la precisión deseada.

Serie de Taylor

Si tomamos en la fórmula (25) un número cada vez mayor de términos, la diferencia entre el segundo miembro y $f(x)$, expresada por el resto $R_{n+1}(x)$, puede tender a cero. Naturalmente, esto no siempre ocurrirá: ni para todas las funciones ni para todos los valores de x . Pero existe una clase muy amplia de funciones (las llamadas funciones *analíticas*) para las cuales el resto $R_{n+1}(x)$ tiende a cero cuando $n \rightarrow \infty$, al menos para todos los valores de x dentro de un cierto intervalo alrededor del punto a . Para estas funciones la fórmula de Taylor permite calcular $f(x)$ con cualquier grado deseado de precisión. Examinémoslas más de cerca.

Si $R_{n+1}(x) \rightarrow 0$ cuando $n \rightarrow \infty$, entonces de (26) se sigue que

$$f(x) = \lim_{n \rightarrow \infty} \left[f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!} (x-a)^n \right].$$

En este caso decimos que hemos desarrollado $f(x)$ en una serie infinita convergente

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!} (x-a)^2 + \cdots,$$

en potencias crecientes de $(x-a)$. Esta serie se llama *serie de Taylor*, y se dice que $f(x)$ es la suma de la serie. Consideremos algunos ejemplos (con $a=0$):

$$1. \quad (1+x)^n = 1 + nx + \frac{n(n-1)}{2!} x^2 + \frac{n(n-1)(n-2)}{3!} x^3 + \cdots$$

(válida para $|x| < 1$ y para un n real cualquiera).

$$2. \quad \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \quad (\text{válido para todo } x).$$

$$3. \quad \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \quad (\text{válido para todo } x).$$

$$4. \quad e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \quad (\text{válido para todo } x).$$

$$5. \quad \arctg x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots \quad (\text{válido para } |x| \leq 1).$$

El primero de estos ejemplos es el famoso teorema del binomio de Newton, quien lo estableció para todo n , pero sólo lo demostró por completo para n entero. Este ejemplo sirvió de modelo para establecer la fórmula general de Taylor. Las dos últimas fórmulas, con $x = 1$, permiten calcular con una aproximación cualquiera los números e y π .

La fórmula de Taylor, que abre el camino para la mayoría de los cálculos en el análisis aplicado, es muy importante desde el punto de vista práctico.

Muchas de las leyes de la naturaleza, procesos físicos y químicos, el movimiento de los cuerpos, etc., se expresan con gran aproximación por funciones que pueden desarrollarse en serie de Taylor. La teoría de tales funciones puede formularse de un modo claro y completo si las consideramos como funciones de variable compleja (ver Cap. 9).

La idea de aproximar una función mediante polinomios o de representarla como suma de un número finito de funciones más sencillas alcanzó un gran desarrollo en el análisis, donde constituye ahora una rama independiente: la teoría de la aproximación de funciones (ver Cap. 12).

§10. Integral

Del capítulo 1 y de la sección 1 del presente capítulo sabemos ya que el concepto de integral —y en general del cálculo integral— tuvo su origen histórico en la necesidad de resolver problemas concretos, uno de cuyos ejemplos más característicos es el cálculo del área de una figura curvilínea. La presente sección está dedicada a estas cuestiones. En ella también se discutirá la conexión entre los problemas del cálculo diferencial e integral, que no quedó completamente aclarada hasta el siglo XVIII.

Area

Supongamos que una curva situada por encima del eje x representa la gráfica de la función $y = f(x)$. Intentemos encontrar el área S de la superficie limitada por la curva $y = f(x)$, el eje x y las rectas que, pasando por los puntos $x = a$ y $x = b$, son paralelas al eje y .

Para resolver este problema se procede como sigue. Dividimos el intervalo $[a, b]$ en n partes, no necesariamente iguales. Notamos

la longitud de la primera parte por Δx_1 , la de la segunda por Δx_2 , y así sucesivamente hasta la última, Δx_n . En cada parte elegimos puntos $\xi_1, \xi_2, \dots, \xi_n$, y escribimos la suma

$$S_n = f(\xi_1) \Delta x_1 + f(\xi_2) \Delta x_2 + \dots + f(\xi_n) \Delta x_n. \quad (28)$$

S_n es evidentemente igual a la suma de las áreas de los rectángulos de la figura 24.

Cuanto más fina sea la subdivisión del segmento $[a, b]$, más próxima se hallará S_n al área S . Si consideramos una sucesión de tales valores por división del intervalo $[a, b]$ en partes cada vez más pequeñas, entonces la suma S_n tenderá a S .

La posibilidad de dividir el intervalo $[a, b]$ en partes desiguales exige definir lo que entendemos por subdivisiones «cada vez más pequeñas». Suponemos no sólo que n crece indefinidamente, sino también que la longitud del

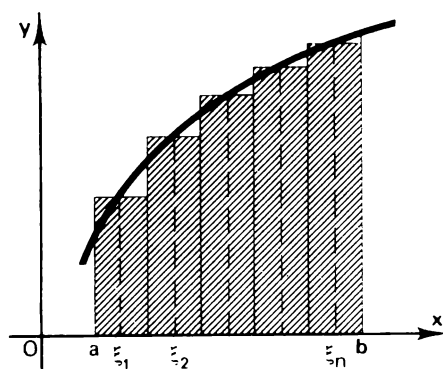


FIG. 24

mayor Δx_i en la n -ésima subdivisión tiende a cero. Así:

$$\begin{aligned} S &= \lim_{\max \Delta x_i \rightarrow 0} [f(\xi_1) \Delta x_1 + f(\xi_2) \Delta x_2 + \dots + f(\xi_n) \Delta x_n] = \\ &= \lim_{\max \Delta x_i \rightarrow 0} \sum_{i=1}^n f(\xi_i) \Delta x_i. \end{aligned} \quad (29)$$

El cálculo del área buscada se ha reducido a calcular el límite (29).

Observemos que al plantear el problema sólo teníamos una idea empírica, y no una definición precisa, de lo que entendíamos por área de la figura curvilínea. Ahora, en cambio, hemos obtenido una definición rigurosa del concepto de área: es el límite (29). Aparte de la noción intuitiva de área, tenemos también una definición matemática, sobre cuya base se puede calcular numéricamente el área

(compárense las observaciones al final del §3, referentes a la velocidad y a la longitud de una circunferencia).

Hemos supuesto que $f(x) \geq 0$. Si $f(x)$ cambia de signo, entonces (en la figura 25) el límite (29) nos dará la suma algebraica de las áreas de las superficies situadas entre la curva $y = f(x)$ y el eje x , tomando con signo $+$ las superficies por encima del eje x , y con signo $-$, las situadas por debajo de él.

Integral definida

La necesidad de calcular el límite (29) aparece en muchos otros problemas. Por ejemplo, supongamos que un cuerpo se mueve a lo largo de una línea recta con velocidad variable $v = f(t)$. ¿Cómo determinaremos la distancia s recorrida por el punto en el intervalo de tiempo entre $t = a$ y $t = b$?

Supongamos que la función $f(t)$ es continua; esto es, que en intervalos pequeños de tiempo la velocidad sólo varía ligeramente. Dividimos el intervalo $[a, b]$ en n partes, de longitudes $\Delta t_1, \Delta t_2, \dots, \Delta t_n$. Para calcular un valor aproximado de la distancia recorrida en cada intervalo Δt_i , supondremos que la velocidad en este intervalo de tiempo es constante e igual a su verdadero valor en algún punto intermedio ξ_i . La distancia total recorrida vendrá expresada entonces aproximadamente por la suma siguiente:

$$s_n = \sum_{i=1}^n f(\xi_i) \Delta t_i,$$

y el verdadero valor de la distancia s recorrida en el tiempo $b - a$ será el límite de tales sumas para subdivisiones más y más finas; esto es, será el límite (29)

$$s = \lim_{\max \Delta t_i \rightarrow 0} \sum_{i=1}^n f(\xi_i) \Delta t_i.$$

Sería fácil poner muchos ejemplos de problemas reales que conducen al cálculo de tales límites. Más tarde analizaremos algunos

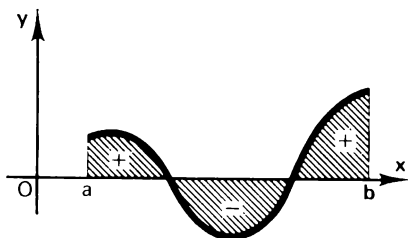


FIG. 25

de ellos, pero por el momento los ejemplos dados hasta aquí son suficientes para hacer comprender la importancia de esta clase de límites. El límite (29) se llama *integral definida* de la función $f(x)$ en el intervalo $[a, b]$, y se nota por

$$\int_a^b f(x) dx.$$

La expresión $f(x) dx$ se llama integrando; a y b son los límites de integración; a es el límite inferior, y b , el límite superior.

Conexión entre el cálculo diferencial e integral

Como ejemplo de cálculo directo de una integral definida podemos considerar el ejemplo 2 de la sección §1. Se ve ahora que el problema allí considerado se reduce al cálculo de la integral definida

$$\int_0^h ax dx.$$

Otro ejemplo, estudiado en §3, consistía en calcular el área limitada por la parábola $y = x^2$. En este caso el problema se reduce al cálculo de la integral siguiente:

$$\int_0^1 x^2 dx.$$

En ambos casos pudimos calcular la integral directamente porque disponíamos de fórmulas sencillas para la suma de los n primeros números naturales y para la suma de sus cuadrados. Pero para una función $f(x)$ cualquiera es prácticamente imposible realizar la suma (28) (es decir, expresar el resultado en una fórmula sencilla) cuando se eligen los puntos ξ_i y los incrementos Δx_i de modo que se ajusten a un problema particular. Además, aun en el caso de que esta suma sea posible, no hay método general para calcularla; sólo en algunos casos pueden emplearse ciertos criterios, que siempre son de carácter muy específico.

Surge así el problema de encontrar un método general para el cálculo de las integrales definidas. Históricamente esta cuestión interesó a los matemáticos durante mucho tiempo, por la utilidad que

ello suponía para el cálculo de áreas de figuras curvilíneas, volúmenes de cuerpos limitados por superficies curvas, etc.

Ya se indicó que Arquímedes logró calcular el área de un segmento de parábola y de otras figuras. El número de problemas particulares que se consiguió resolver (cálculo de áreas, volúmenes, centros de gravedad de sólidos, etc.) fue creciendo gradualmente, pero los progresos en lo referente a encontrar un método general fueron, al principio, extremadamente lentos. Dicho método sólo fue descubierto cuando se hubo acumulado suficiente material teórico y práctico, proporcionado por la experiencia diaria. El trabajo de recoger y generalizar este material avanzó muy lentamente hasta el final de la Edad Media; y su rápido desarrollo posterior fue una consecuencia directa del acelerado crecimiento del poder productivo de Europa como resultado de la desaparición de los primitivos métodos (feudales) de producción, y la aparición de otros nuevos (capitalistas).

La acumulación de datos relacionados con las integrales definidas marchó paralela a la investigación de problemas relacionados con la derivada de una función. El lector ya sabe —por lo dicho en la sección §1— que esta inmensa labor preparatoria fue coronada con éxito en el siglo XVIII por los trabajos de Newton y Leibnitz. En este sentido se puede decir que Newton y Leibnitz son los creadores del cálculo diferencial e integral.

Una de las contribuciones fundamentales de Newton y Leibnitz fue que aclararon finalmente la profunda conexión entre el cálculo diferencial e integral, que proporciona, en particular, un método general para calcular las integrales definidas de una clase bastante amplia de funciones.

Para aclarar esta conexión, analicemos un ejemplo tomado de la mecánica.

Supongamos que un punto material se mueve a lo largo de una línea recta con velocidad $v = f(t)$, donde t es el tiempo. Se sabe que la distancia σ recorrida por el punto en el intervalo del tiempo desde $t = t_1$ a $t = t_2$ viene expresada por la integral definida

$$\sigma = \int_{t_1}^{t_2} f(t) dt.$$

Supongamos conocida la ecuación del movimiento del punto material; esto es, la función $s = F(t)$, que expresa la dependencia de la distancia s respecto del tiempo t a partir de un punto inicial A sobre

la recta. La distancia σ recorrida en el intervalo de tiempo $[t_1, t_2]$ es evidentemente igual a la diferencia

$$\sigma = F(t_2) - F(t_1).$$

De este modo, consideraciones físicas llevan a la igualdad

$$\int_{t_1}^{t_2} f(t) dt = F(t_2) - F(t_1),$$

que expresa la conexión entre la ecuación del movimiento del punto material y su velocidad.

Desde un punto de vista matemático la función $F(t)$ —como ya sabemos por lo dicho en la sección §5— puede definirse como una función cuya derivada, para todos los valores de t del intervalo dado, es igual a $f(t)$, esto es:

$$F'(t) = f(t).$$

Una tal función se llama *primitiva* de $f(t)$.

Hay que tener en cuenta que si la función $f(t)$ tiene al menos una primitiva, entonces tiene un número infinito de ellas; porque si $F(t)$ es una primitiva de $f(t)$, entonces $F(t) + C$ (donde C es una constante arbitraria) es también una primitiva. Además, de este modo obtenemos todas las primitivas de $f(t)$, puesto que si $F_1(t)$ y $F_2(t)$ son primitivas de la misma función $f(t)$, entonces su diferencia, $\phi(t) = F_1(t) - F_2(t)$, tiene una derivada $\phi'(t)$ que es igual a cero en todo punto del intervalo dado, por lo que $\phi(t)$ es constante²⁴.

Desde un punto de vista físico los diferentes valores de la constante C determinan movimientos que sólo difieren entre sí en el hecho de que corresponden a todas las posibles elecciones del punto inicial del movimiento.

Se llega así al resultado de que para una clase muy amplia de funciones $f(x)$, que incluye todos los casos en los que la función $f(x)$

²⁴ Por el teorema del valor medio

$$\phi(t) - \phi(t_0) = \phi'(v)(t - t_0) = 0,$$

donde v se encuentra entre t y t_0 . Así $\phi(t) = \phi(t_0) = \text{constante}$ para todo t .

pueda ser considerada como velocidad de un punto en el instante x , se verifica la siguiente igualdad ²⁵:

$$\int_a^b f(x) dx = F(b) - F(a), \quad (30)$$

donde $F(x)$ es una primitiva cualquiera de $f(x)$.

Esta igualdad es la famosa *fórmula de Newton y Leibnitz*, que reduce el problema de calcular la integral definida de una función a la obtención de una primitiva de la misma, y constituye así un enlace entre el cálculo diferencial y el integral.

Muchos de los problemas concretos estudiados por los más grandes matemáticos se resuelven automáticamente con esta fórmula, que establece sencillamente que la integral definida de la función $f(x)$ en el intervalo $[a, b]$ es igual a la diferencia entre los valores de cualquiera de sus primitivas en los extremos superior e inferior del intervalo ²⁶. La diferencia (30) se acostumbra a escribir así:

$$F(x) \Big|_a^b = F(b) - F(a).$$

Ejemplo 1

La igualdad

$$\left(\frac{x^3}{3} \right)' = x^2$$

muestra que la función $x^3/3$ es una primitiva de la función x^2 . Así, por la fórmula de Newton y Leibnitz,

$$\int_0^a x^2 dx = \frac{x^3}{3} \Big|_0^a = \frac{a^3}{3} - \frac{0}{3} = \frac{a^3}{3}.$$

²⁵ Es posible demostrar matemáticamente, sin recurrir a ejemplos de la mecánica, que si la función $f(x)$ es continua (o incluso si es discontinua pero sumable Lebesgue; ver capítulo 15) sobre el intervalo $[a, b]$, entonces existe una primitiva $F(x)$ que satisface la igualdad (30).

²⁶ Esta fórmula ha sido generalizada en distintas formas (ver por ejemplo §13, la fórmula de Ostrogradski).

Ejemplo 2

Sean c y c' dos cargas eléctricas, situadas sobre una línea recta y a distancia r una de la otra. La atracción F entre ellas está dirigida a lo largo de esta recta y es igual a

$$F = \frac{a}{r^2}$$

($a = kcc'$, donde k es una constante). El trabajo W realizado por esta fuerza, cuando la carga c permanece fija y la c' se mueve a lo largo del intervalo $[R_1, R_2]$ puede calcularse dividiendo el intervalo $[R_1, R_2]$ en partes Δr_i . Sobre cada una de estas partes podemos considerar la fuerza como aproximadamente constante, por lo que el trabajo realizado sobre cada parte es igual a $(a/r_i^2) \Delta r_i$. Tomando divisiones cada vez más pequeñas, vemos que el trabajo W es igual a la integral

$$W = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{a}{r_i^2} \Delta r_i = \int_{R_1}^{R_2} \frac{a}{r^2} dr.$$

El valor de esta integral puede calcularse rápidamente, si recordamos que

$$\frac{a}{r^2} = \left(-\frac{a}{r} \right)',$$

por lo que

$$W = -\frac{a}{r} \Big|_{R_1}^{R_2} = a \left(\frac{1}{R_1} - \frac{1}{R_2} \right).$$

En particular, el trabajo realizado por una fuerza F cuando la carga c , inicialmente a la distancia R_1 de c , se aleja hasta el infinito, es igual a

$$W = \lim_{R_2 \rightarrow \infty} a \left(\frac{1}{R_1} - \frac{1}{R_2} \right) = \frac{a}{R_1}.$$

De los razonamientos hechos al exponer la fórmula de Newton y Leibnitz se desprende claramente que esta fórmula es la expresión matemática a una conexión auténtica del mundo real. Es un bello e importante ejemplo de cómo la matemática da expresión a las leyes objetivas. Debemos observar que, en sus investigaciones matemáticas,

Newton siempre adoptó un punto de vista físico. Sus trabajos sobre los fundamentos del cálculo diferencial e integral no pueden ser separados de sus trabajos sobre los principios de la mecánica.

Los conceptos del análisis matemático —como la derivada o la integral— tal como se presentaban a Newton y sus contemporáneos, aún no habían «roto» del todo con sus orígenes físico y geométrico (velocidad y área). De hecho, eran de un carácter mitad matemático y mitad físico. Las condiciones existentes en esa época no eran todavía las apropiadas para lograr una definición puramente matemática de estos conceptos. Por consiguiente, el investigador sólo podía manejarlos correctamente en situaciones complejas si permanecía en contacto inmediato con los aspectos prácticos del problema incluso durante las etapas intermedias (matemáticas) de su razonamiento.

Desde este punto de vista el trabajo creador de Newton tuvo un carácter diferente del de Leibnitz²⁷. Newton se dejó guiar siempre por el enfoque físico de los problemas. En cambio las investigaciones de Leibnitz no tienen una conexión tan inmediata con la física, hecho que, en ausencia de definiciones matemáticas precisas, le condujo a veces a conclusiones equivocadas. Por otra parte, el rasgo más característico de la actividad creadora de Leibnitz fue su esfuerzo por generalizar, su búsqueda de los métodos más generales de resolución de los problemas del análisis matemático.

El mayor mérito de Leibnitz fue la creación de un simbolismo matemático que expresaba lo esencial de la cuestión. Las notaciones para conceptos fundamentales del análisis matemático, tales como la diferencial dx , la diferencial segunda d^2x , la integral $\int y dx$, y la derivada d/dx fueron propuestas por Leibnitz. El hecho de que estas notaciones se utilicen todavía muestra lo acertado de su elección.

Una de las ventajas de un simbolismo bien elegido es que hace las demostraciones y cálculos más cortos y fáciles, y evita también, a veces, conclusiones equivocadas. Leibnitz, quien no ignoraba esto, prestó especial atención en todo su trabajo a la elección de notaciones.

La evolución de los conceptos del análisis matemático (derivada, integral, etc.) continuó, naturalmente, después de Newton y Leibnitz y continúa todavía en nuestros días; pero hay una etapa en esta evolución que merece ser destacada. Tuvo lugar a comienzos del siglo pasado y está particularmente relacionada con el trabajo de Cauchy.

²⁷ Los descubrimientos de Newton y Leibnitz tuvieron lugar independientemente.

Cauchy dio una definición formal precisa del concepto de límite y la utilizó como base para sus definiciones de continuidad, derivada, diferencial e integral. Estos conceptos, que han sido presentados ya en los correspondientes lugares de este capítulo, se emplean constantemente en el análisis moderno.

La gran importancia de estos resultados reside en el hecho de que gracias a ellos es posible operar de un modo puramente formal y llegar a conclusiones correctas no sólo en la aritmética, el álgebra y la geometría elemental, sino también en esa nueva y extensa rama de la matemática, el análisis matemático.

En cuanto a los resultados prácticos del análisis matemático, se puede decir hoy lo siguiente: si los datos originales se toman del mundo real, entonces el resultado de los razonamientos matemáticos también se verificará en él. Y, si estamos completamente seguros de la precisión de los datos originales, no hay necesidad de hacer una comprobación práctica de la exactitud de los resultados matemáticos; basta comprobar la exactitud de los razonamientos formales.

Esta afirmación tiene naturalmente la siguiente limitación. En los razonamientos matemáticos los datos originales, que tomamos del mundo real, sólo son verdaderos con una cierta precisión. Esto significa que en cada etapa del razonamiento matemático los resultados obtenidos contendrán ciertos errores que, conforme avanza el razonamiento²⁸, pueden irse acumulando.

Volviendo ahora a la integral definida, consideremos una cuestión de capital importancia. ¿Para qué funciones $f(x)$, definidas sobre el intervalo $[a, b]$, es posible garantizar la existencia de la integral definida $\int_a^b f(x) dx$, es decir, un número para el cual la suma $\sum_1^n f(\xi_i) \Delta x_i$ tenga límite cuando $\max \Delta x_i \rightarrow 0$? Debe tenerse en cuenta que este número será el mismo para todas las subdivisiones del intervalo $[a, b]$ y para cualquier elección de los puntos ξ_i .

Las funciones para las cuales la integral definida, es decir, el límite (29), existe, se dicen *integrables* en el intervalo $[a, b]$. Investigaciones realizadas en el último siglo demostraron que todas las funciones continuas son integrables.

Pero hay también funciones discontinuas que son integrables, y entre ellas figuran por ejemplo las funciones que son acotadas y crecientes (o decrecientes) en el intervalo $[a, b]$.

²⁸ Por ejemplo, de $a = b$ y $b = c$ se sigue formalmente que $a = c$. Pero en la práctica esta relación aparece como sigue: del hecho de que $a = b$ con un error igual a ϵ , y $b = c$ con un error también ϵ , se sigue que $a = c$ con un error igual a 2ϵ .

La función que es igual a cero en los puntos racionales de $[a, b]$ e igual a uno en los puntos irracionales puede servir de ejemplo de función no integrable, puesto que para una subdivisión arbitraria la suma s_n será igual a cero o a uno, según que elijamos los puntos ξ_i entre los números irracionales o los racionales.

Señalemos que en muchos casos la fórmula de Newton y Leibnitz proporciona la solución al problema de calcular una integral definida. Pero surge entonces el problema de encontrar una primitiva de una función dada; esto es, de encontrar una función que tenga por derivada la función dada. Procedamos ahora a discutir este punto. Observemos de paso que el problema de encontrar una primitiva tiene gran importancia también en otras ramas de la matemática, particularmente en la solución de ecuaciones diferenciales.

§11. Integrales indefinidas; técnica de integración

Una primitiva arbitraria de una función dada $f(x)$ se denomina generalmente *integral indefinida* de $f(x)$ y se escribe en la forma

$$\int f(x) dx.$$

De este modo, si $F(x)$ es una primitiva perfectamente determinada de $f(x)$, la integral indefinida de $f(x)$ viene dada por

$$\int f(x) dx = F(x) + C, \quad (31)$$

donde C es una constante arbitraria.

Notemos también que, si la función $f(x)$ está definida en el intervalo $[a, b]$, $F(x)$ es una primitiva de $f(x)$ y x es un punto del intervalo $[a, b]$, entonces por la fórmula de Newton y Leibnitz podemos escribir

$$F(x) = F(a) + \int_a^x f(t) dt.$$

En esta igualdad la integral del segundo miembro difiere sólo de la primitiva $F(x)$ en la constante $F(a)$. En tal caso dicha integral, si la consideramos como función de su límite superior x (para x

variable), es una primitiva perfectamente definida de $f(x)$. Por consiguiente, una integral indefinida de $f(x)$ puede también escribirse así:

$$\int f(x) dx = \int_a^x f(t) dt + C,$$

donde C es una constante arbitraria.

He aquí una tabla fundamental de integrales indefinidas que puede obtenerse directamente de la correspondiente tabla de derivadas (véase §6):

$$\left. \begin{aligned} \int x^a dx &= \frac{x^{a+1}}{a+1} + C \quad (a \neq -1), \\ \int \frac{dx}{x} &= \ln |x| + C^{29}, \\ \int a^x dx &= \frac{a^x}{\ln a} + C, \\ \int e^x dx &= e^x + C, \\ \int \operatorname{sen} x dx &= -\cos x + C, \\ \int \cos x dx &= \operatorname{sen} x + C, \\ \int \sec^2 x dx &= \operatorname{tg} x + C, \\ \int \frac{dx}{\sqrt{1-x^2}} &= \operatorname{arc} \operatorname{sen} x + C \\ &= -\operatorname{arc} \cos x + C_1 \left(C_1 - C = \frac{\pi}{2} \right), \\ \int \frac{dx}{1+x^2} &= \operatorname{arc} \operatorname{tg} x + C. \end{aligned} \right\} \quad (32)$$

²⁹ Para $x > 0$, $(\ln |x|)' = (\ln x)' = 1/x$; para $x < 0$, $(\ln |x|)' = (\ln |-x|)' = 1/|-x|(-1) = 1/x$.

Las propiedades generales de las integrales indefinidas pueden también deducirse de las propiedades homólogas de las derivadas. Por ejemplo, de la regla de derivación de una suma obtenemos la fórmula

$$\int [f(x) \pm \phi(x)] dx = \int f(x) dx \pm \int \phi(x) dx + C;$$

y de la regla que expresa que un factor constante k puede extraerse fuera del signo de derivación, obtenemos:

$$\int k f(x) dx = k \int f(x) dx + C.$$

Por ejemplo,

$$\begin{aligned} \int \left(3x^2 + 2x - \frac{3}{\sqrt{x}} + \frac{4}{x} - 1 \right) dx &= \\ &= 3 \frac{x^3}{3} + 2 \frac{x^2}{2} - 3 \frac{x^{-1/2+1}}{-1/2+1} + 4 \ln |x| - x + C. \end{aligned}$$

Hay distintos métodos para calcular integrales indefinidas. Consideremos uno de ellos, el llamado *método de sustitución* o de cambio de variable, que se basa en la igualdad siguiente:

$$\int f(x) dx = \int f[\phi(t)] \phi'(t) dt + C, \quad (33)$$

donde $x = \phi(t)$ es una función diferenciable. La relación (33) se debe entender en el sentido de que, si en la función

$$F(x) = \int f(x) dx,$$

del primer miembro de la igualdad (33) hacemos $x = \phi(t)$, obtenemos una función $F[\phi(t)]$ cuya derivada respecto a t es igual a la expresión bajo el signo integral en el segundo miembro de la igualdad (33). Esto se deduce inmediatamente del teorema sobre derivación de una función de función.

Veamos algunos ejemplos de este método de sustitución:

$$\int e^{kx} dx = \int e^t \frac{1}{k} dt = \frac{1}{k} \int e^t dt = \frac{1}{k} e^t + C = \frac{e^{kx}}{k} + C$$

(sustitución $kx = t$, $k dx = dt$).

$$\int \frac{x dx}{\sqrt{a^2 - x^2}} = - \int dt = -t + C = -\sqrt{a^2 - x^2} + C$$

(sustitución $t = \sqrt{a^2 - x^2}$, $dt = -\frac{x dx}{\sqrt{a^2 - x^2}}$).

$$\begin{aligned} \int \sqrt{a^2 - x^2} dx &= \int \sqrt{a^2 - a^2 \sin^2 u} a \cos u du = a^2 \int \cos^2 u du = \\ &= a^2 \int \frac{1 + \cos 2u}{2} du = \frac{a^2}{2} \left(u + \frac{\sin 2u}{2} \right) + C = \\ &= \frac{a^2}{2} (u + \sin u \cos u) + C = \\ &= \frac{a^2}{2} \left(\arcsin \frac{x}{a} + \frac{x}{a^2} \sqrt{a^2 - x^2} \right) + C \end{aligned}$$

(sustitución $x = a \sin u$).

Como se ve en estos ejemplos, el método de sustitución o de cambio de variable amplía bastante la clase de funciones elementales que el lector es ahora capaz de integrar; esto es, aquellas para las que pueden encontrarse primitivas que a su vez sean funciones elementales. Pero debe observarse que, desde el punto de vista del cálculo efectivo del resultado, nuestra situación es mucho peor, en general, respecto a la integración que respecto a la derivación.

Desde el §6 sabemos que la derivada de una función elemental es a su vez una función elemental que puede efectivamente calcularse utilizando las reglas de derivación. Pero la afirmación recíproca es en general falsa, puesto que existen funciones elementales cuyas integrales indefinidas no son funciones elementales. Ejemplos son: e^{-x^2} , $1/(\ln x)$, $(\sin x)/x$, etc. Para obtener la integral de estas funcio-

nes es preciso utilizar métodos aproximados e introducir también nuevas funciones que no pueden ser reducidas a las elementales. No podemos profundizar más aquí sobre esta cuestión, pero queremos señalar que incluso en la matemática elemental es posible encontrar muchos ejemplos de operaciones que se puedan efectuar sobre una cierta clase de números, mientras que la operación inversa no puede realizarse sobre esa misma clase; así, el cuadrado de un número racional es también un número racional, y sin embargo la raíz cuadrada de un número racional no siempre es racional. Análogamente, la derivada de las funciones elementales es una función también elemental, mientras que la integración puede conducirnos fuera de la clase de las funciones elementales.

Algunas de las integrales que no pueden expresarse en términos de funciones elementales tienen gran importancia en la matemática y en sus aplicaciones. Un ejemplo es

$$\int_0^x e^{t^{-2}} dt,$$

integral muy importante en la teoría de probabilidades (véase capítulo 11). Otros ejemplos son las integrales

$$\int_0^\phi \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \quad \text{y} \quad \int_0^\phi \sqrt{1 - k^2 \sin^2 \theta} \, d\theta \quad (k^2 < 1),$$

que se llaman *integrales elípticas* de primera y segunda especie, respectivamente. Un gran número de problemas de la física conducen al cálculo de estas integrales (ver capítulo 5, §1, ejemplo 3). Utilizando métodos aproximados se han calculado con gran exactitud tablas detalladas de estas integrales para diversos valores de los argumentos x y ϕ .

La demostración del simple hecho de que la integral de una función elemental dada no puede expresarse en términos de funciones elementales es, en cada caso concreto, bastante difícil. Tales cuestiones ocuparon la atención de relevantes matemáticos del último siglo y han sido parte importante en el desarrollo del análisis. Chebichev obtuvo en este campo resultados fundamentales, dando completa respuesta a la cuestión de expresar en términos de funciones elementales las integrales de la forma

$$\int x^m (a + bx^s)^p dx,$$

donde m , s y p son números racionales. Antes de su época se conocían tres relaciones —obtenidas por Newton— entre los exponentes m , s y p , para las cuales la integral se podía expresar en términos de las funciones elementales. Chebichev demostró que en todos los demás casos la integral no puede ser expresada mediante funciones elementales.

Introduzcamos ahora otro método de integración, el método de integración por partes. Se basa en la conocida fórmula

$$(uv)' = uv' + u'v,$$

para la derivada del producto de las funciones u y v . Esta fórmula puede también escribirse

$$uv' = (uv)' - u'v.$$

Integremos ambos miembros de la igualdad, teniendo en cuenta que

$$\int (uv)' dx = uv + C.$$

Obtenemos así finalmente la igualdad

$$\int uv' dx = uv - \int u'v dx,$$

que se llama *fórmula de integración por partes*. No hemos escrito la constante C , puesto que podemos considerar que va incluida en una de las integrales indefinidas que aparecen en esta ecuación.

Veamos algunas aplicaciones de esta fórmula. Supongamos que tenemos que calcular $\int xe^x dx$. Hagamos $u = x$ y $v' = e^x$, con lo cual $u' = 1$, $v = e^x$, y por consiguiente:

$$\int xe^x dx = xe^x - \int 1 \cdot e^x dx = xe^x - e^x + C.$$

En la integral $\int \ln x dx$ es conveniente tomar $u = \ln x$, $v' = 1$, por lo que $u' = 1/x$, $v = x$ y

$$\int \ln x dx = x \ln x - \int dx = x \ln x - x + C.$$

En el siguiente ejemplo característico es necesario integrar dos veces por partes y luego despejar la integral buscada:

$$\begin{aligned}\int e^x \operatorname{sen} x \, dx &= e^x \operatorname{sen} x - \int e^x \cos x \, dx = \\ &= e^x \operatorname{sen} x - e^x \cos x - \int e^x \operatorname{sen} x \, dx,\end{aligned}$$

de la cual

$$\int e^x \operatorname{sen} x \, dx = \frac{e^x}{2} (\operatorname{sen} x - \cos x) + C.$$

Con ello cerramos esta sección, de la cual el lector habrá obtenido sólo una idea superficial de la teoría de la integración. No hemos prestado atención a otros muchos métodos de esta teoría. En particular no hemos mencionado la muy interesante cuestión de la integración de fracciones racionales, teoría a la que el conocido matemático y físico Ostrogradski tanto aportó.

§12. *Funciones de varias variables*

Hasta ahora sólo hemos hablado de funciones de una variable, pero en la práctica es necesario a menudo tratar también con funciones que dependen de dos, tres, o en general varias variables. Por ejemplo, el área de un rectángulo es una función

$$S = xy$$

de su base x y su altura y . El volumen de un paralelepípedo rectangular es una función

$$v = xyz$$

de sus tres dimensiones. La distancia entre dos puntos A y B es una función

$$r = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

de las seis coordenadas de estos puntos. La conocida fórmula

$$pv = RT$$

expresa la relación entre el volumen v de una cantidad dada de gas y la presión p y temperatura absoluta T .

En muchos casos, las funciones de varias variables —como las de una variable— están definidas únicamente sobre una cierta región de valores de las variables. Por ejemplo, la función

$$u = \ln(1 - x^2 - y^2 - z^2) \quad (34)$$

sólo está definida para los valores de x , y y z que satisfacen la condición

$$x^2 + y^2 + z^2 < 1. \quad (35)$$

(Para otros valores de x , y , z , la función no toma valores reales.) El conjunto de puntos del espacio cuyas coordenadas satisfacen la desigualdad (35) son, evidentemente, los de la esfera de radio unidad y centro en el origen de coordenadas. Los puntos del contorno no están incluidos en esta esfera: una superficie «mondada», cabría llamarla. Una tal esfera se dice que es abierta. La función (34) está definida únicamente para las ternas de números (x, y, z) que son coordenadas de los puntos de la esfera abierta G . Se acostumbra expresar esto concisamente diciendo que la función (34) está definida sobre la esfera G .

Veamos otro ejemplo. La temperatura de un cuerpo V calentado de un modo no uniforme es función de las coordenadas x, y, z de los puntos del cuerpo. Esta función no está definida para todas las ternas de números (x, y, z) , sino sólo para aquellas que son coordenadas de puntos del cuerpo V .

Finalmente, como tercer ejemplo, consideremos la función

$$u = \phi(x) + \phi(y) + \phi(z),$$

donde ϕ es función de una variable definida en el intervalo $[0, 1]$. Evidentemente, la función u sólo está definida para ternas de números (x, y, z) que sean coordenadas de puntos del cubo:

$$0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad 0 \leq z \leq 1.$$

Veamos ahora una definición formal de función de tres variables. Supongamos dado un conjunto E de ternas de números (x, y, z) (puntos del espacio). Si a cada una de estas ternas (puntos) de E le corresponde un número u de acuerdo con cierta ley, se dice que u es

una función de x, y, z (del punto), definida sobre el conjunto de ternas de números (sobre los puntos) de E , hecho que se puede escribir así:

$$u = F(x, y, z).$$

En lugar de F podemos utilizar también otras letras: f, ϕ, ψ .

En la práctica F es en general un conjunto de puntos que llena por completo un cuerpo o superficie geométrica: esfera, cubo, anillo, etcétera, y se dice sencillamente que la función está definida sobre este cuerpo o superficie. De manera análoga se definen las funciones que dependen de dos, cuatro, etc., variables.

Definición implícita de una función

Es de notar que las funciones de dos variables pueden servir, en ciertas circunstancias, como medio útil para la definición de funciones de una variable. Dada una función $F(x, y)$ de dos variables, formemos la ecuación

$$F(x, y) = 0. \quad (36)$$

En general, esta ecuación definirá un cierto conjunto de puntos (x, y) de la superficie sobre la cual la función es igual a cero. Tales conjuntos de puntos representan generalmente curvas que pueden considerarse como gráficas de una o varias funciones uniformes de una variable $y = \phi(x)$ o $x = \psi(y)$. En tal caso estas funciones uniformes se dice que están definidas implícitamente por la ecuación (36). Por ejemplo, la ecuación

$$x^2 + y^2 - r^2 = 0$$

da una definición implícita de dos funciones de una variable:

$$y = +\sqrt{r^2 - x^2} \text{ e } y = -\sqrt{r^2 - x^2}.$$

Pero conviene tener en cuenta que una ecuación de la forma (36) puede no definir ninguna función. Por ejemplo, la ecuación

$$x^2 + y^2 + 1 = 0$$

no define evidentemente ninguna función real, puesto que ningún par de números reales la satisface.

Representación geométrica

Las funciones de dos variables pueden siempre visualizarse como superficies por medio de un sistema de coordenadas espaciales. Así, la función

$$z = f(x, y) \quad (37)$$

representa en un sistema de coordenadas rectangulares tridimensional una superficie, que es el lugar geométrico de los puntos M cuyas coordenadas x, y, z satisfacen la ecuación (37) (Fig. 26).

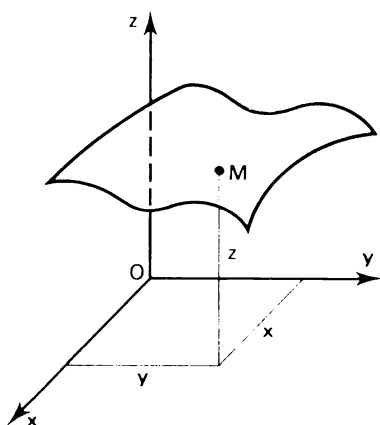


FIG. 26

Hay otro método, extremadamente útil, de representar la función (37), que ha encontrado gran aplicación en la práctica. Elija-mos una sucesión de números $z_1, z_2, \dots$, y luego dibujemos sobre el mismo plano Oxy las curvas

$$z_1 = f(x, y), \quad z_2 = f(x, y),$$

que se llaman curvas de nivel de la función $f(x, y)$. Con un conjunto de curvas de nivel correspondiente a valores de z suficientemente próximos, es posible formarse una imagen muy aproximada de la variación de la función $f(x, y)$, del mismo modo que

se hace para juzgar la variación en altitud de una localidad a partir de las curvas de nivel de un mapa topográfico.

La figura 27 muestra un mapa de las curvas de nivel de la función $z = x^2 + y^2$, indicando el diagrama de la derecha cómo se puede obtener la función mediante dichas curvas. En el capítulo 3, figura 50, se ha dibujado un mapa similar para las curvas de nivel de la función $z = xy$.

Derivadas parciales y diferencial

Hagamos algunas observaciones sobre la derivación de funciones de varias variables. Como ejemplo consideremos una función de dos variables

$$z = f(x, y).$$

Si fijamos el valor de y , esto es, si consideramos que y no varía, la función de dos variables se convierte en función de la única variable x . La derivada de esta función con respecto a x , si existe, se llama *derivada parcial* con respecto a x , y se nota así:

$$\frac{\partial z}{\partial x}, \quad \text{o} \quad \frac{\partial f}{\partial x}, \quad \text{o} \quad f'_x(x, y).$$

La última de las tres notaciones indica claramente que la derivada

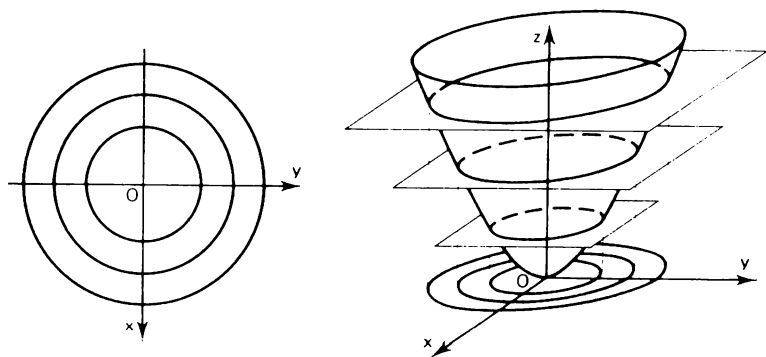


FIG. 27

parcial con respecto a x es, en general, una función de x e y . La derivada parcial con respecto a y se define de modo similar.

Geométricamente la función $f(x, y)$ representa una superficie en un sistema tridimensional de coordenadas rectangular. La correspondiente función de x , para y fijo, representa una curva plana (Fig. 28), obtenida por intersección de la superficie con un plano paralelo al plano Oxz a una distancia y de él. La derivada parcial $\partial z/\partial x$ es evidentemente igual a la tangente trigonométrica del ángulo formado por la tangente a la curva en el punto (x, y) y la dirección positiva del eje x .

En general, si consideramos una función $z = f(x_1, x_2, \dots, x_n)$ de las n variables $x_1, x_2, \dots, x_n$, la derivada parcial $\partial z/\partial x_i$ se define como la derivada de esta función respecto a x_i , calculada para valores fijos de las otras variables:

$$x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n.$$

Se puede decir que la derivada parcial de una función con respecto a la variable x_i es el ritmo de variación de esta función en la dirección del eje x_i . También es posible definir una derivada en una dirección cualquiera dada, no necesariamente coincidente con uno de los ejes coordenados, aunque no lo haremos aquí.

Ejemplos

$$1. \quad z = \frac{x}{y}; \quad \frac{\partial z}{\partial x} = \frac{1}{y}; \quad \frac{\partial z}{\partial y} = -\frac{x}{y^2}.$$

$$2. \quad u = \frac{1}{\sqrt{x^2 + y^2 + z^2}};$$

$$\frac{\partial u}{\partial x} = -\frac{1}{x^2 + y^2 + z^2} \cdot \frac{2x}{2\sqrt{x^2 + y^2 + z^2}} = -\frac{x}{(x^2 + y^2 + z^2)^{3/2}}.$$

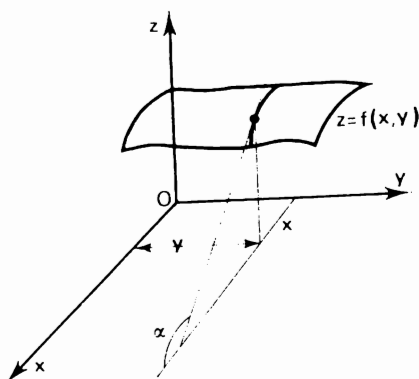


FIG. 28

A veces es necesario hallar las derivadas parciales de estas derivadas parciales; esto es, las llamadas derivadas parciales de segundo orden. Para funciones de dos variables hay cuatro de ellas:

$$\frac{\partial^2 u}{\partial x^2}, \quad \frac{\partial^2 u}{\partial x \partial y}, \quad \frac{\partial^2 u}{\partial y \partial x}, \quad \frac{\partial^2 u}{\partial y^2}.$$

Sin embargo, si estas derivadas son continuas, no es difícil probar que la segunda y tercera (las llamadas derivadas cruzadas) coinciden:

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x}$$

Por ejemplo, en el caso de la primera función considerada,

$$\frac{\partial^2 z}{\partial x^2} = 0, \quad \frac{\partial^2 z}{\partial x \partial y} = -\frac{1}{y^2}, \quad \frac{\partial^2 z}{\partial y \partial x} = -\frac{1}{y^2}, \quad \frac{\partial^2 z}{\partial y^2} = \frac{2x}{y^3}$$

donde se ve que las dos derivadas cruzadas coinciden.

Para funciones de varias variables, al igual que se hizo para las de una variable, se puede definir el concepto de diferencial.

Para precisar consideremos una función de dos variables:

$$z = f(x, y).$$

Si tiene derivadas parciales continuas, se puede demostrar que su incremento

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y),$$

correspondiente a los incrementos Δx y Δy de sus argumentos, puede escribirse en la forma

$$\Delta z = \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y + \alpha \sqrt{\Delta x^2 + \Delta y^2},$$

donde $\partial f/\partial x$ y $\partial f/\partial y$ son las derivadas parciales de la función en el punto (x, y) , y la variable α depende de Δx e Δy de modo tal que $\alpha \rightarrow 0$ cuando $\Delta x \rightarrow 0$ e $\Delta y \rightarrow 0$.

La suma de los dos primeros términos

$$dz = \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y$$

depende linealmente³⁰ de Δx e Δy , y se llama *diferencial de la función*. El tercer sumando, debido a la presencia del factor α que tiende a cero con Δx e Δy , es un infinitésimo de orden superior al

$$\rho = \sqrt{\Delta x^2 + \Delta y^2},$$

que determina la variación en x e y .

Veamos una aplicación del concepto de diferencial. El periodo de oscilación de un péndulo se calcula por la fórmula

$$T = 2\pi \sqrt{\frac{l}{g}},$$

³⁰ En general una función $Ax + By + C$, donde A, B, C son constantes, se dice función lineal de x e y . Si $C = 0$, se llama función lineal homogénea. Aquí se omitirá la palabra «homogénea».

donde l es su longitud y g la aceleración de la gravedad. Supongamos que l y g se conocen con errores respectivamente iguales a Δl e Δg . Entonces el error en T será igual al incremento ΔT , correspondiente a los incrementos Δl e Δg de las magnitudes l y g . Reemplazando ΔT aproximadamente por dT , tendremos:

$$\Delta T \approx dT = \pi \left(\frac{\Delta l}{\sqrt{lg}} - \frac{\sqrt{l} \Delta g}{\sqrt{g^3}} \right).$$

Los signos de Δl e Δg no se conocen, pero podemos evidentemente estimar ΔT por la desigualdad

$$|\Delta T| \leq \pi \left(\frac{|\Delta l|}{\sqrt{lg}} + \sqrt{\frac{l}{g^3}} |\Delta g| \right);$$

dividiendo por T , obtenemos

$$\frac{|\Delta T|}{T} < \left(\frac{|\Delta l|}{l} + \frac{|\Delta g|}{g} \right).$$

Así, podemos considerar prácticamente que el error relativo de T es igual a la suma de los errores relativos de l y g .

Por simetría en la notación, los incrementos de las variables independientes Δx e Δy se notan usualmente con los símbolos dx y dy , y se llaman también diferenciales. Con esta notación la diferencial de la función $u = f(x, y, z)$ puede escribirse así:

$$du = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

Las derivadas parciales son importantes en todo lo que se refiera a funciones de varias variables, como sucede en muchas de las aplicaciones del análisis a la tecnología y a la física. En el capítulo 6 estudiaremos el problema de reconstruir una función a partir de las propiedades de sus derivadas parciales.

En las secciones siguientes veremos algunos sencillos ejemplos de aplicación de las derivadas parciales al análisis.

Derivación de las funciones implícitas

Supongamos que deseamos encontrar la derivada de y , donde y es función de x definida implícitamente por la relación

$$F(x, y) = 0 \quad (38)$$

entre estas variables. Si x e y satisfacen la relación (38) y damos a x el incremento Δx , entonces y se incrementará en Δy de tal modo que $x + \Delta x$ e $y + \Delta y$ satisfagan también (38). Por tanto ³¹:

$$F(x + \Delta x, y + \Delta y) - F(x, y) = \frac{\partial F}{\partial x} \Delta x + \frac{\partial F}{\partial y} \Delta y + \alpha \sqrt{\Delta x^2 + \Delta y^2} = 0.$$

Así, siempre que $\partial F / \partial y \neq 0$, se sigue que

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = y'_x = - \frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}.$$

De este modo hemos encontrado un método para calcular la derivada de una función implícita y sin despejar previamente la y en la ecuación (38).

Problemas de máximos y mínimos

Si una función, por ejemplo de dos variables $z = f(x, y)$, tiene un máximo en el punto (x_0, y_0) , esto es, si $f(x_0, y_0) \geq f(x, y)$ para todos los puntos (x, y) próximos a (x_0, y_0) , entonces este punto debe ser también punto máximo para cualquier curva que sea intersección de la superficie $z = f(x, y)$ con un plano paralelo al Oxz o al Oyz . Así en este punto debemos tener

$$f'_x(x, y) = 0, \quad f'_y(x, y) = 0. \quad (39)$$

Las mismas ecuaciones deben también verificarse para un mínimo relativo. Por consiguiente, para hallar los máximos y mínimos de la función se buscan primero los puntos que satisfagan las condiciones (39), aunque sin olvidar los puntos del contorno del dominio de

³¹ Supondremos que $F(x, y)$ tiene derivadas continuas respecto a x e y .

definición de la función, ni los puntos en los que la función no tenga derivada, si es que existen dichos puntos.

Para establecer si un punto (x, y) que satisface (39) es realmente un punto máximo o mínimo, se hace uso con frecuencia de diversos razonamientos indirectos. Por ejemplo, si por cualquier razón se sabe que la función es derivable, que alcanza su mínimo dentro de la región y que sólo hay un punto en el que se cumplen las condiciones (39), entonces evidentemente el mínimo debe alcanzarse en este punto.

Por ejemplo, se pide construir una caja de hojalata rectangular (sin tapadera) de volumen dado V , empleando la menor cantidad posible de material. Si denotamos por x e y los lados de la base de esta caja, entonces su altura h será igual a V/xy , y por tanto la superficie S vendrá dada por la función de x e y siguiente:

$$S = xy + \frac{V}{xy}(2x + 2y) = xy + 2V\left(\frac{1}{x} + \frac{1}{y}\right) \quad (40)$$

Puesto que x e y , por las condiciones del problema, deben ser positivos, el problema se reduce a obtener el mínimo de la función $S(x, y)$ para todos los puntos posibles (x, y) del primer cuadrante del plano (x, y) , que denotaremos con la letra G .

Si se alcanza el mínimo en algún punto de la región G , entonces las derivadas parciales deben ser iguales a cero:

$$\frac{\partial S}{\partial x} = y - \frac{2V}{x^2} = 0$$

$$\frac{\partial S}{\partial y} = x - \frac{2V}{y^2} = 0,$$

esto es, $yx^2 = 2V$, $xy^2 = 2V$, por lo que las dimensiones de la caja son:

$$x = y = \sqrt[3]{2V} \quad y \quad h = \sqrt[3]{\frac{V}{4}}. \quad (41)$$

Hemos resuelto el problema, pero no hemos demostrado todavía que esta solución es la correcta. Un matemático diría: «Usted ha supuesto desde el principio que bajo las condiciones dadas existe realmente una caja de superficie mínima y, a partir de esta suposición,

ha encontrado sus dimensiones. Así que realmente sólo ha obtenido usted el siguiente resultado: si existe un punto (x, y) de G para el cual la función S alcanza su mínimo, entonces las coordenadas de este punto deben estar necesariamente determinadas por la ecuación (41). Pero ahora debe demostrar que el mínimo de S existe para algún punto de G , y entonces yo admitiré que su resultado es correcto.» Esta observación es muy razonable, puesto que, la función S , como pronto veremos, no tiene máximo en la región G . Pero demostremos ahora cómo es posible convencernos de que en este caso la función alcanza realmente su mínimo en un cierto punto (x, y) de la región G .

El teorema fundamental sobre el cual basaremos nuestros razonamientos se demuestra con todo rigor en análisis. Dice lo siguiente: si la función f de una o varias variables es continua en todo punto de una región finita H que está acotada e incluye su contorno, existe siempre en H un punto al menos en el que la función alcanza su mínimo (máximo). Con este teorema podemos fácilmente completar nuestro análisis del problema.

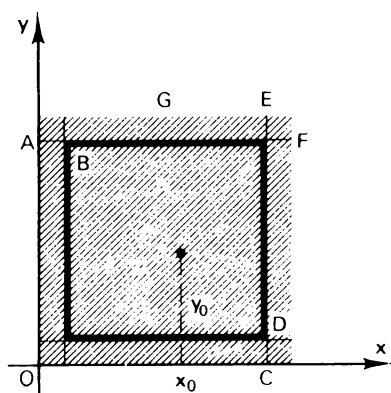


FIG. 29

Consideremos un punto arbitrario (x_0, y_0) de la región G ; sea $S(x_0, y_0) = N$ en este punto. Elijamos también un número R que satisfaga las dos desigualdades $R > N$, $2VR > N$, y construyamos un cuadrado Ω_R de lado R^2 (como en la figura 29), donde $AB = CD = 1/R$.

Busquemos ahora una cota inferior para los valores de la función $S(x, y)$ en los puntos de la región G situados fuera del cuadrado Ω_R . Si el punto de la región G tiene abscisa $x < 1/R$, entonces

$$S(x, y) = xy + 2V \left(\frac{1}{x} + \frac{1}{y} \right) > 2V \frac{1}{x} > 2VR > N.$$

Análogamente, si la ordenada del punto de la región G cumple $y < 1/R$, entonces también $S > N$. Además, si el punto de la región G

tiene abscisa $x > 1/R$ y se encuentra por encima de la recta AF , o tiene su ordenada $y > 1/R$ y se encuentra a la derecha de la recta CE , entonces

$$S(x, y) > xy > \frac{1}{R} R^2 = R > N.$$

Así, para todos los puntos (x, y) de la región G situados fuera del cuadrado Ω_R , se verifica la desigualdad $S(x, y) > N$; y, puesto que $S(x_0, y_0) = N$, el punto (x_0, y_0) debe pertenecer al cuadrado, por lo cual el mínimo de la función en G es igual a su mínimo en el cuadrado.

Pero la función $S(x, y)$ es continua en este cuadrado y en su frontera, por lo que en virtud del teorema establecido anteriormente existirá en el cuadrado un punto (x, y) en el que la función tendrá un mínimo para los puntos del cuadrado y, por tanto, para toda la región G . Queda demostrada así la existencia de un mínimo.

Este razonamiento puede servir como ejemplo del modo de discutir la existencia de un máximo o mínimo para una función definida sobre un dominio no acotado.

Fórmula de Taylor

Al igual que las funciones de una variable, las funciones de varias variables pueden representarse por una fórmula de Taylor. Por ejemplo, el desarrollo de la función

$$u = f(x, y)$$

en el entorno del punto (x_0, y_0) tiene la siguiente forma si nos limitamos a las primeras y segundas potencias de $x - x_0$ e $y - y_0$:

$$\begin{aligned} f(x, y) = & f(x_0, y_0) + [f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0)] + \\ & + \frac{1}{2!} [f''_{xx}(x_0, y_0)(x - x_0)^2 + 2f''_{xy}(x_0, y_0)(x - x_0)(y - y_0) + \\ & + f''_{yy}(x_0, y_0)(y - y_0)^2] + R_3. \end{aligned}$$

Si la función $f(x, y)$ tiene derivadas parciales continuas de segundo orden, el resto tenderá a cero más rápidamente que

$$r^2 = (x - x_0)^2 + (y - y_0)^2,$$

esto es, más rápidamente que el cuadrado de la distancia entre los puntos (x, y) y (x_0, y_0) , cuando $r \rightarrow 0$. La fórmula de Taylor constituye un método muy útil de definir y calcular aproximadamente valores de funciones.

Observemos que con la ayuda de esta fórmula podemos también resolver la cuestión (antes planteada) de si una función dada tiene realmente máximo o mínimo en un punto donde $\partial f / \partial x = \partial f / \partial y = 0$. En efecto, si estas condiciones se satisfacen en el punto (x_0, y_0) , entonces para los puntos (x, y) próximos a (x_0, y_0) el valor de la función diferirá de $f(x_0, y_0)$, de acuerdo con la fórmula de Taylor, en la cantidad

$$f(x, y) - f(x_0, y_0) = \frac{1}{2!} [A(x - x_0)^2 + 2B(x - x_0)(y - y_0) + C(y - y_0)^2] + R_3, \quad (42)$$

donde A, B y C notan respectivamente a las derivadas parciales segundas $f''_{xx}, f''_{xy}, f''_{yy}$ en el punto (x_0, y_0) .

Si se comprueba que la función

$$\Phi(x, y) = A(x - x_0)^2 + 2B(x - x_0)(y - y_0) + C(y - y_0)^2$$

es positiva para cualesquiera valores de $(x - x_0)$ e $(y - y_0)$ no simultáneamente nulos, entonces el segundo miembro de la ecuación (42) será también positivo para valores pequeños de $(x - x_0)$ e $(y - y_0)$, puesto que para $(x - x_0)$ e $(y - y_0)$ suficientemente pequeños R_3 será menor, en valor absoluto, que $\Phi(x, y)$. Se deduce así que en el punto (x_0, y_0) la función f alcanza su mínimo. Por otra parte, si la función $\Phi(x, y)$ es negativa para cualesquiera $(x - x_0)$ e $(y - y_0)$, el segundo miembro de (42) será negativo para $(x - x_0)$ e $(y - y_0)$, por lo que en el punto (x_0, y_0) la función tendrá un máximo. En casos más complicados es necesario considerar un número mayor de términos en la fórmula de Taylor.

Los problemas que se refieren a máximos y mínimos de funciones de tres o más variables pueden tratarse de manera completamente análoga. Como ejercicio puede demostrar el lector que dadas las masas

$$m_1, m_2, \dots, m_n$$

colocadas en puntos fijos del espacio

$$P_1(x_1, y_1, z_1), P_2(x_2, y_2, z_2), \dots, P_n(x_n, y_n, z_n),$$

el momento (de inercia) M de este sistema de masas respecto al punto $P(x, y, z)$, definido como la suma de los productos de las masas por los cuadrados de sus distancias al punto P ,

$$M(x, y, z) = \sum_{i=1}^n m_i [(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2],$$

es mínimo cuando el punto P es el llamado centro de gravedad del sistema de coordenadas

$$x = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}, \quad y = \frac{\sum_{i=1}^n m_i y_i}{\sum_{i=1}^n m_i}, \quad z = \frac{\sum_{i=1}^n m_i z_i}{\sum_{i=1}^n m_i}.$$

Máximos y mínimos condicionados

Para las funciones de varias variables podemos plantear diversos problemas sobre máximos y mínimos. Ilustrémoslo con un sencillo ejemplo. Supongamos que entre todos los rectángulos inscritos en

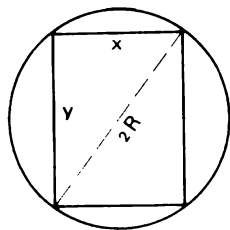


FIG. 30

un círculo de radio R deseamos encontrar el de mayor área. El área de un rectángulo es igual al producto xy de sus lados, donde x e y son números positivos relacionados en este caso por la condición $x^2 + y^2 = (2R)^2$, como se deduce claramente de la figura 30. Así, pues, se pide encontrar el máximo de la función $f(x, y) = xy$ donde los puntos x e y satisfacen la condición $x^2 + y^2 = 4R^2$.

Problemas de esta clase, en los que es necesario encontrar el máximo (o mínimo) de una función $f(x, y)$ para valores de x e y que satisfagan una cierta relación del tipo $\phi(x, y) = 0$ son muy comunes en la práctica.

Naturalmente, cabría despejar y en la ecuación $\phi(x, y) = 0$, sustituir la solución en la función $f(x, y)$, y de este modo buscar el máximo de una función de una sola variable x . Pero este método es generalmente complicado, y a veces, imposible.

Para resolver estos problemas se ha desarrollado un procedimiento analítico, mucho más conveniente, llamado método de los

multiplicadores de Lagrange. La idea en la que se funda es extremadamente sencilla. Consideremos la función

$$F(x, y) = f(x, y) + \lambda \phi(x, y),$$

donde λ es un número positivo arbitrario. Evidentemente, para x e y tales que satisfagan la condición $\phi(x, y) = 0$, los valores de $F(x, y)$ coincidirán con los de $f(x, y)$.

Busquemos los máximos de la función $F(x, y)$ considerando las variables x e y como independientes. En el punto de máximo deben cumplirse las condiciones $\partial F / \partial x = \partial F / \partial y = 0$ ³²; en otras palabras:

$$\frac{\partial f}{\partial x} + \lambda \frac{\partial \phi}{\partial x} = 0; \quad (43)$$

$$\frac{\partial f}{\partial y} + \lambda \frac{\partial \phi}{\partial y} = 0. \quad (44)$$

Los valores de x e y correspondientes al punto de máximo de $F(x, y)$, al ser solución del sistema (43) y (44), dependen del coeficiente λ de estas ecuaciones. Supongamos ahora que hemos logrado encontrar un número λ tal que las coordenadas del punto de máximo satisfacen la condición

$$\phi(x, y) = 0. \quad (45)$$

Este punto será entonces un verdadero máximo relativo en el problema original.

De hecho, el problema se puede considerar geométricamente como sigue. Supongamos que la función $f(x, y)$ está definida en una cierta región G (Fig. 31). La condición $\phi(x, y) = 0$ será satisfecha ordinariamente por los puntos de cierta curva Γ . Se pide encontrar el mayor valor de $f(x, y)$ para los puntos de la línea Γ . Si $F(x, y)$ alcanza el máximo sobre la curva Γ , entonces $F(x, y)$ no crece para pequeños incrementos en una dirección arbitraria a partir de este punto, y en particular para incrementos a lo largo de la curva Γ . Pero para los incrementos a lo largo de Γ , los valores de $F(x, y)$ coinciden con los de $f(x, y)$, lo cual significa que, para pequeños desplazamientos según la curva, la función $f(x, y)$ no crece, o en otras palabras, tiene un máximo relativo en ese punto.

³² Hablamos aquí, naturalmente, del máximo alcanzado en el dominio de definición de la función $F(x, y)$. Las funciones $f(x, y)$ y $\phi(x, y)$ se suponen diferenciables.

Estos razonamientos ilustran un método sencillo de solucionar el problema. Se resuelven las ecuaciones (43), (44) y (45) para las incógnitas x , y y λ , obteniéndose una o más soluciones

$$(x_1, y_1, \lambda_1), (x_2, y_2, \lambda_2), \dots \quad (46)$$

A los puntos (x_1, y_1) , (x_2, y_2) , ... así determinados añadimos aquellos puntos del contorno de G en los que la curva Γ corta a la región G . De todos estos puntos elegimos aquel para el cual $f(x, y)$ toma su valor máximo o mínimo.

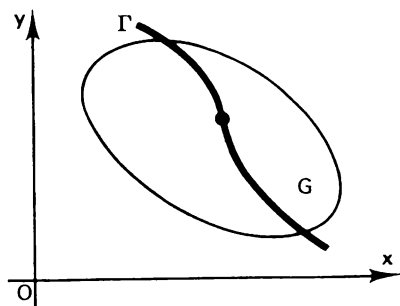


FIG. 31

Naturalmente, los razonamientos expuestos hasta aquí no justifican la validez del método. De hecho, ni siquiera hemos demostrado todavía que los máximos relativos de $f(x, y)$ sobre la curva Γ puedan obtenerse como máximos de la función $F(x, y)$ para algún valor de λ . Sin embargo, es posible demostrar, como se hace

en los textos de análisis, que todo punto (x_0, y_0) de la curva en el que $f(x, y)$ tiene un máximo relativo se puede obtener por el método indicado, con tal, únicamente, de que en este punto las derivadas parciales $\phi'_x(x_0, y_0)$ y $\phi'_y(x_0, y_0)$ no sean simultáneamente nulas³³.

Utilicemos el método de Lagrange para resolver el problema planteado al comienzo de la sección. En este caso es $f(x, y) = xy$; $\phi(x, y) = x^2 + y^2 - 4R^2$. Planteamos las ecuaciones (43), (44), (45)

$$\begin{aligned} y + 2\lambda x &= 0, \\ x + 2\lambda y &= 0, \\ x^2 + y^2 &= 4R^2, \end{aligned}$$

³³ En el curso de matemática superior de V. I. Smirnov, el lector encontrará un sencillo ejemplo en donde esta situación particular conduciría a la pérdida de una solución si se aplica mecánicamente el método de Lagrange y no se consideran, junto con los puntos mencionados anteriormente, los que no sólo verifiquen (45) sino también:

$$\phi'_x(x_0, y_0) = 0, \quad \phi'_y(x_0, y_0) = 0.$$

de las cuales (teniendo en cuenta que x e y son positivos) encontramos la solución única

$$x = y = R\sqrt{2} \left(\lambda = -\frac{1}{2} \right).$$

Para estos valores de x e y (iguales, por lo cual el rectángulo inscrito es un cuadrado) el área es máxima.

El método de Lagrange puede extenderse a funciones de tres o más variables. El número de condiciones puede ser entonces arbitrario (aunque menor que el número de variables). Las condiciones serán del tipo de la condición (45), y se introducirá el correspondiente número de multiplicadores auxiliares.

Veamos algunos ejemplos de problemas que envuelven máximos o mínimos condicionados.

Ejemplo 1

¿Para qué altura h y radio r un tanque cilíndrico abierto de volumen dado V requerirá la menor cantidad de chapa metálica para su construcción? Esto es, ¿cuándo la suma del área lateral y de la base circular será mínima?

El problema evidentemente se reduce a encontrar el mínimo de la siguiente función de las variables r y h :

$$f(r, h) = 2\pi rh + \pi r^2,$$

con la condición $\pi r^2 h = V$, que puede escribirse en la forma

$$\phi(r, h) = \pi r^2 h - V = 0.$$

Ejemplo 2

Se desea que un punto en movimiento pase de A a B (Fig. 32). En el trayecto AM se mueve con la velocidad v_1 , y en el MB con la velocidad v_2 . ¿Dónde debería situarse el punto M sobre la línea DD' para que el trayecto de A a B sea recorrido lo más rápido posible?

Tomemos como incógnitas los ángulos α y β señalados en la

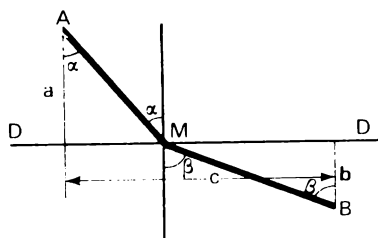


FIG. 32

figura 32. Las longitudes a y b de las perpendiculares desde los puntos A y B a la recta DD' y la distancia c entre ellas son conocidas. El tiempo necesario para recorrer todo el camino viene expresado —como puede deducirse fácilmente— por la fórmula

$$f(\alpha, \beta) = \frac{a}{v_1 \cos \alpha} + \frac{b}{v_2 \cos \beta}.$$

Se pide encontrar el mínimo de esta expresión, teniendo en cuenta el hecho de que α y β están relacionados por la ecuación

$$a \operatorname{tg} \alpha + b \operatorname{tg} \beta = c.$$

El lector puede resolver estos ejemplos por el método de Lagrange. En el segundo ejemplo encontrará que la posición de M viene dada por la condición

$$\frac{\operatorname{sen} \alpha}{\operatorname{sen} \beta} = \frac{v_1}{v_2}.$$

Esta es la conocida ley para la refracción de la luz. Por consiguiente, al pasar de un medio a otro el rayo de luz se refractará de modo tal que el tiempo de paso entre dos puntos fijos situados en distintos medios será mínimo. Conclusiones de esta clase no sólo son interesantes para el cálculo, sino también desde un punto de vista filosófico general; ellas han inspirado a los científicos en esa labor de penetrar más y más en las leyes, profundas y generales, de la naturaleza.

Observemos que los multiplicadores λ , introducidos para la resolución de problemas por el método de Lagrange, no son simplemente números auxiliares. En todos los casos están íntimamente relacionados con la naturaleza esencial de cada problema particular y tienen una interpretación concreta.

§13. Generalizaciones del concepto de integral

En el §10 definimos la integral definida de la función $f(x)$ en el intervalo $[a, b]$ como el límite de la suma

$$\sum_{i=1}^n f(\xi_i) \Delta x_i$$

cuando la longitud del mayor segmento Δx_i en la subdivisión de $[a, b]$ tiende a cero. A pesar de que la clase de funciones $f(x)$ para las cuales existe realmente este límite (la clase de las funciones integrables) es muy amplia —en particular incluye todas las funciones continuas, e incluso muchas discontinuas—, dicha clase presenta un grave inconveniente. Si se suman, restan, multiplican o, en ciertas condiciones, dividen dos funciones integrables $f(x)$ y $\phi(x)$, se obtienen funciones que, como puede demostrarse fácilmente, son también integrables. Para $f(x)/\phi(x)$ esto será cierto siempre que $1/\phi(x)$ permanezca acotado en $[a, b]$. Pero si la función se obtiene como resultado de un proceso de límite de una sucesión de funciones integrables $f_1(x), f_2(x), f_3(x), \dots$, tales que para cada uno de los valores de x del intervalo $[a, b]$

$$f(x) = \lim_{n \rightarrow \infty} f_n(x),$$

entonces la función límite $f(x)$ no es necesariamente integrable.

En muchos casos, esta circunstancia da lugar a complicaciones considerables, puesto que el proceso de paso al límite se utiliza con mucha frecuencia.

Una salida a esta dificultad consistió en generalizar el concepto de integral. La generalización más importante de todas es la integral de Lebesgue, que el lector encontrará en el capítulo 14 como parte de la teoría de funciones de variable real. Pero aquí nos limitaremos a otras generalizaciones de la integral, que son también de máxima importancia en la práctica.

Integrales múltiples

Ya hemos estudiado el proceso de integración para funciones de una variable, definidas sobre una región unidimensional, es decir, sobre un intervalo. De modo análogo puede ser extendido este concepto a funciones de dos, tres o más variables, definidas sobre regiones correspondientes.

Consideremos, por ejemplo, la superficie

$$z = f(x, y)$$

definida en un sistema de coordenadas rectangular, y supongamos que nos dan en el plano Oxy una región G limitada por una curva cerrada Γ . Se pide encontrar el volumen encerrado o limitado por la superficie, el plano Oxy y la superficie cilíndrica que pasa por la

curva I' con generatrices paralelas al eje Oz (Fig. 33). Para resolver este problema se divide la región plana G en subregiones, mediante un cuadrículado de lados paralelos a los ejes Ox y Oy , notando por

$$G_1, G_2, \dots, G_n$$

aquellas subregiones que constan de rectángulos completos. Si la red es suficientemente fina, prácticamente toda la región G quedará recubierta por los rectángulos completos. En cada uno de ellos se elige un punto

$$(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_n, \eta_n),$$

y suponiendo, para simplificar, que G_i nota no sólo el rectángulo, sino también su área, escribimos la suma

$$S_n = f(\xi_1, \eta_1) G_1 + f(\xi_2, \eta_2) G_2 + \dots + f(\xi_n, \eta_n) G_n = \sum_{i=1}^n f(\xi_i, \eta_i) G_i. \quad (47)$$

Está claro que, si la superficie es continua y la red suficientemente fina, esta suma puede aproximarse tanto como deseemos al volumen pedido V . La magnitud exacta de este volumen se obtiene calculando el límite de la suma (47) para subdivisiones cada vez más finas (esto es, para subdivisiones tales que la mayor de las dos diagonales de los rectángulos tienda a cero):

$$\lim_{\max d(G_i) \rightarrow 0} \sum_{i=1}^n f(\xi_i, \eta_i) G_i = V. \quad (48)$$

Así, pues, para determinar el volumen V es preciso, desde el punto de vista del análisis, realizar una cierta operación ma-

temática sobre la función $f(x, y)$ y su dominio de definición G , operación que viene indicada por el primer miembro de la igualdad (48). Esta operación se llama integración de la función f sobre la

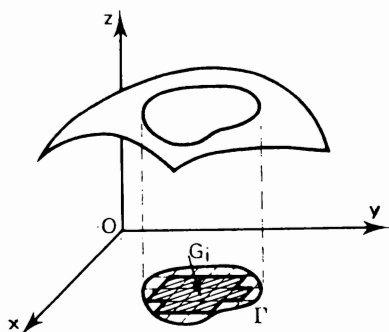


FIG. 33

región G , y su resultado es la integral de f sobre G . Se acostumbra a notar este resultado del modo siguiente:

$$\iint_G f(x, y) dx dy = \lim_{\max d(G_i) \rightarrow 0} \sum_{i=1}^n f(\xi_i, \eta_i) G_i. \quad (49)$$

De modo similar se define la integral de una función de tres variables sobre una región tridimensional G , representada por un cierto cuerpo en el espacio. De nuevo dividimos la región G en partes, esta vez por medio de planos paralelos a los planos coordenados. Entre estas partes se eligen aquellas que representan paralelepípedos completos y los enumeramos

$$G_1, G_2, \dots, G_n.$$

En cada uno de éstos elegimos un punto arbitrario

$$(\xi_1, \eta_1, \zeta_1), (\xi_2, \eta_2, \zeta_2), \dots, (\xi_n, \eta_n, \zeta_n)$$

y calculamos la suma

$$S = \sum_{i=1}^n f(\xi_i, \eta_i, \zeta_i) G_i, \quad (50)$$

donde G_i nota el volumen del paralelepípedo G_i . Finalmente se define la integral de $f(x, y, z)$ sobre la región G como el límite siguiente:

$$\lim_{\max d(G_i) \rightarrow 0} \sum_{i=1}^n f(\xi_i, \eta_i, \zeta_i) G_i = \iiint_G f(x, y, z) dx dy dz, \quad (51)$$

al cual tiende la suma (50) cuando la diagonal mayor $d(G_i)$ tiende a cero.

Veamos un ejemplo. Imaginemos que la región G está ocupada por una masa no homogénea cuya densidad, en cada punto de G , viene dada por una función conocida $\varrho(x, y, z)$. La densidad $\varrho(x, y, z)$ en el punto (x, y, z) se define como el límite al que tiende el cociente entre la masa de una región arbitrariamente pequeña que contiene al punto (x, y, z) y el volumen de esta región cuando su diámetro tiende

a cero ³⁴. Para determinar la masa del cuerpo G es natural proceder como sigue. Dividimos la región G mediante planos paralelos a los planos coordenados y enumeramos los paralelepípedos completos así obtenidos del modo siguiente:

$$G_1, G_2, \dots, G_n.$$

Suponiendo que los planos divisores estén suficientemente próximos unos a otros, el error que cometeremos al despreciar las regiones irregulares del cuerpo será pequeño. Así, pues, podemos definir la masa de cada una de las regiones regulares G_i (los paralelepípedos completos) como el producto

$$\varrho(\xi_i, \eta_i, \zeta_i) G_i,$$

donde (ξ_i, η_i, ζ_i) es un punto arbitrario de G_i . Por consiguiente, el valor aproximado de la masa M vendrá expresado por la suma

$$S = \sum_{i=1}^n \varrho(\xi_i, \eta_i, \zeta_i) G_i,$$

y su valor exacto será claramente el límite de esta suma cuando la diagonal mayor de los G_i tienda a cero; esto es,

$$M = \iiint_G \varrho(x, y, z) dx dy dz = \lim_{\max d(G_i) \rightarrow 0} \sum_{i=1}^n \varrho(\xi_i, \eta_i, \zeta_i) G_i.$$

Las integrales (49) y (51) se llaman integrales doble y triple, respectivamente.

Analicemos un problema que conduce a una integral doble. Imaginemos una superficie plana sobre la cual fluye agua. La intensidad $f(x, y)$ con que se filtra el agua —hacia arriba o hacia abajo— es diferente en puntos distintos. Consideremos una región G limitada por un contorno cerrado (Fig. 34) y supongamos que en todo punto de G conocemos la intensidad $f(x, y)$, es decir, la cantidad de agua que se filtra por minuto y por cm^2 de superficie; tendremos $f(x, y) > 0$ allí donde el terreno absorba el agua subterránea, y $f(x, y) < 0$ donde

³⁴ El diámetro de una región se define como la mínima cota superior de la distancia entre dos puntos de la región.

el agua se filtre hacia abajo. ¿Cuánta agua se acumulará sobre la superficie G por minuto?

Si dividimos G en partes pequeñas, consideramos la velocidad de filtración como aproximadamente constante en cada parte y pasamos luego al límite para subdivisiones cada vez más finas, obtendremos para la cantidad total de agua acumulada una expresión del tipo siguiente:

$$\iint_G f(x, y) dx dy.$$

Las integrales dobles fueron introducidas por primera vez por Euler. Las integrales múltiples constituyen un instrumento que se utiliza a diario en cálculos e investigaciones de los más variados tipos.

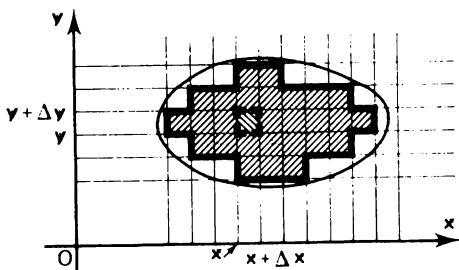


FIG. 34

Es también posible demostrar (aunque no lo haremos aquí) que el cálculo de las integrales múltiples puede reducirse, en general, al cálculo reiterado de integrales ordinarias unidimensionales.

Integrales de superficie y de línea

Finalmente, debemos mencionar algunas otras generalizaciones de la integral. Por ejemplo, el problema de definir el trabajo realizado por una fuerza variable aplicada a un punto material, cuando éste se mueve a lo largo de una curva dada, conduce de manera natural a las llamadas integrales curvilíneas; y el problema de encontrar la carga eléctrica sobre una superficie sobre la cual dicha carga se distribuye de forma continua con una densidad superficial dada, conduce a otro nuevo concepto, el de integral sobre una superficie curva.

Por ejemplo, supongamos un líquido en movimiento (Fig. 35) en el que la velocidad de la partícula que se halla en el punto (x, y) viene dada por una función $P(x, y)$ que no depende de z . Si se desea calcular la cantidad de líquido que fluye por minuto a través del contorno Γ^{35} , podemos razonar del modo siguiente. Dividamos Γ en segmentos Δs_i .

³⁵ Precisando: a través de una superficie cilíndrica que tenga el contorno por base y altura igual a uno.

La cantidad de líquido que fluye a través del segmento Δs_i es aproximadamente igual a la columna de líquido sombreada en la figura 35; esta columna puede considerarse como la cantidad de líquido que atraviesa esa parte del contorno por minuto. Pero el área del paralelogramo sombreado es igual a

$$P_i(x, y) \cdot \Delta s_i \cdot \cos \alpha_i,$$

donde α_i es el ángulo formado por la dirección positiva $\vec{x}$ del eje x y la normal exterior a la superficie limitada por el contorno Γ ; esta

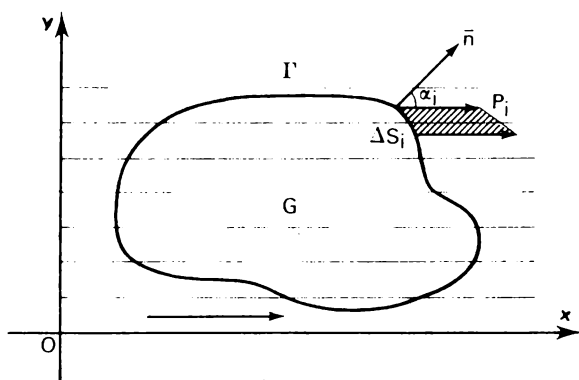


FIG. 35

normal es la perpendicular $\vec{n}$ a la tangente, que puede considerarse que define la dirección del segmento Δs_i . Sumando las áreas de tales paralelogramos y pasando al límite para subdivisiones más y más finas del contorno Γ , se determina la

cantidad de agua que fluye por minuto a través del mismo; se nota así:

$$\int_{\Gamma} P(x, y) \cos(\vec{n}, \vec{x}) ds$$

y se llama integral curvilínea. Si el flujo no es paralelo en todos los puntos, la velocidad en cada punto (x, y) tendrá una componente $P(x, y)$ a lo largo del eje x y una componente $Q(x, y)$ a lo largo del eje y . En este caso se puede demostrar por un razonamiento análogo que la cantidad de agua que atraviesa el contorno será igual a

$$\int_{\Gamma} [P(x, y) \cos(\vec{n}, \vec{x}) + Q(x, y) \cos(\vec{n}, \vec{y})] ds. \quad ^{36}$$

³⁶ Puesto que para pequeños desplazamientos a lo largo de la curva la dife-

Cuando hablamos de la integral sobre una superficie curva G de una función $f(M)$, donde $M(x, y, z)$ son los puntos de G , nos referimos al límite de sumas de la forma

$$\lim \sum_{i=1}^n f(M_i) \Delta\sigma_i = \iint_G f(x, y, z) d\sigma$$

para subdivisiones cada vez más finas de la región G en segmentos de área igual a $\Delta\sigma_i$.

Existen métodos generales para pasar de un tipo de integral a otro (múltiples, curvilíneas y de superficie) y para calcular sus valores o aproximados.

Fórmula de Ostrogradski

A mediados del siglo pasado Ostrogradski encontró varias fórmulas importantes y muy generales que permiten pasar de una integral de volumen a una integral de superficie (y también de una integral de superficie —curva o plana— a una integral sobre su contorno).

No trataremos de dar aquí una demostración de la fórmula general de Ostrogradski, que tiene muchas aplicaciones, sino que nos limitaremos a ilustrarla con un ejemplo referente a su caso particular más sencillo.

Imaginemos, como antes, que sobre una superficie plana corre un flujo horizontal de agua, y que este flujo puede absorber o perder líquido a través de la superficie. Delimitemos una región G , contenida dentro de la curva I , y supongamos que para cada punto de la región conocemos las componentes $P(x, y)$ y $Q(x, y)$ de la velocidad del agua en la dirección de los ejes x e y , respectivamente.

Calculemos la velocidad con que el flujo absorbe agua a través de la superficie en un punto de coordenadas (x, y) . Para ello consideremos un pequeño rectángulo de lados Δx e Δy situados en el punto (x, y) .

Como resultado de la velocidad $P(x, y)$, a través del lado vertical izquierdo de este rectángulo fluirán *hacia dentro* aproximadamente $P(x, y)\Delta y$ unidades de agua por minuto; y a través del lado derecho fluirán *hacia afuera* aproximadamente $P(x + \Delta x, y)\Delta y$ unidades en

rencial de la coordenada y es igual a $\cos(\vec{n}, \vec{x}) ds$ y la diferencial dx es igual a $-\cos(\vec{n}, \vec{y}) ds$, esta última integral se escribe a menudo en la forma

$$\int_I [P(x, y) dy - Q(x, y) dx].$$

el mismo tiempo. Así, pues, la cantidad neta de agua que abandona la unidad cuadrada de superficie como resultado del flujo a través de los lados verticales izquierdo y derecho será aproximadamente

$$\frac{[P(x + \Delta x, y) - P(x, y)] \Delta y}{\Delta x \Delta y}$$

Si hacemos tender Δx a cero, obtenemos en el límite

$$\frac{\partial P}{\partial x}.$$

Igualmente, la velocidad neta de flujo de agua por unidad de área en la dirección del eje y vendrá dada aproximadamente por

$$\frac{\partial Q}{\partial y}.$$

Esto significa que la intensidad de filtración en el punto de coordenadas (x, y) será igual a

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}.$$

Pero en general, como vimos antes, la cantidad de agua que entra en el flujo viene dada por la integral doble de la función que expresa la intensidad de filtración de agua en cada punto, es decir:

$$\iint_G \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} \right) dx dy. \quad (52)$$

Ahora bien, puesto que el agua es incompresible, esta misma cantidad debe fluir hacia afuera en el mismo tiempo a través del contorno Γ . La cantidad de agua que sale por el contorno Γ se expresa, como vimos antes, por la integral curvilínea sobre Γ :

$$\int_{\Gamma} [P(x, y) \cos(\vec{n}, \vec{x}) + Q(x, y) \cos(\vec{n}, \vec{y})] ds. \quad (53)$$

El resultado de igualar las expresiones (52) y (53) constituye la fórmula de Ostrogradski en su caso bidimensional más sencillo:

$$\iint_G \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} \right) dx dy = \int_r [P(x, y) \cos(\vec{n}, \vec{x}) + Q(x, y) \cos(\vec{n}, \vec{y})] ds.$$

Aunque hemos explicado el significado de esta fórmula valiéndonos de un ejemplo físico, puede ser demostrada matemáticamente.

De este modo el teorema matemático de Ostrogradski refleja un fenómeno muy corriente en el mundo físico y que, en el ejemplo anterior, corresponde a la invariancia del volumen de un fluido incompresible.

Ostrogradski estableció una fórmula mucho más general que expresa la conexión entre la integral sobre un volumen multidimensional y la integral sobre la superficie del mismo. En particular, para un cuerpo tridimensional G , limitado por la superficie Γ , la fórmula es

$$\begin{aligned} \iiint_G \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dx dy dz &= \\ &= \iint_{\Gamma} [P \cos(\vec{n}, \vec{x}) + Q \cos(\vec{n}, \vec{y}) + R \cos(\vec{n}, \vec{z})] d\sigma, \end{aligned}$$

donde $d\sigma$ es el elemento de superficie.

Es interesante observar que la fórmula fundamental del cálculo integral

$$\int_a^b f(x) dx = F(b) - F(a) \quad (54)$$

puede considerarse como el caso unidimensional de la fórmula de Ostrogradski. La ecuación (54) relaciona la integral sobre un intervalo con la «integral» sobre su contorno «de dimensión nula», que consiste en sus dos puntos extremos.

La fórmula (54) puede ilustrarse por medio de la siguiente analogía. Imaginemos que a través de una tubería recta de sección transversal constante $s = 1$ fluye agua con una velocidad $F(x)$ que depende de la sección transversal (Fig. 36). A través de las paredes porosas

de la tubería fluye agua (hacia adentro o hacia afuera) a una velocidad que depende también de la sección transversal.

Si consideramos el segmento de tubería comprendido entre x y $x + \Delta x$, la cantidad de agua que penetra en él por unidad de tiempo debe compensarse con la diferencia $F(x + \Delta x) - F(x)$ entre la cantidad que sale del segmento y la que entra por la tubería. Así,

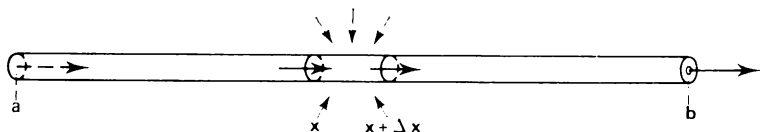


FIG. 36

la cantidad que entra en el segmento es igual a la diferencia $F(x + \Delta x) - F(x)$, y consiguientemente la velocidad de filtración por unidad de longitud de tubería (cociente de la filtración en una parte de tubería infinitesimal entre la longitud de la parte) será igual a

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x).$$

En general, la cantidad de agua que se filtra en la tubería a lo largo de toda la sección $[a, b]$ debe ser igual a la cantidad perdida por flujo a través de los terminales de la tubería. Pero la cantidad absorbida a través de las paredes es igual a $\int_a^b f(x) dx$ y la cantidad perdida por los extremos es $F(b) - F(a)$. La igualación de estos dos valores constituye la fórmula (54).

§14. Series

Concepto de serie

En matemáticas, una serie es una expresión de la forma

$$u_0 + u_1 + u_2 + \dots$$

Los números u_k se llaman términos de la serie. Hay un número infinito de ellos, y están colocados en un orden dado, por lo que a todo número natural $k = 0, 1, 2, \dots$ corresponde un término definido u_k .

Obsérvese que no hemos dicho si es posible o no asignar un valor a tales expresiones y, caso de ser posible, cómo hacerlo. La presencia de un signo de sumar entre los términos u_k de la expresión parece indicar que de algún modo todos los términos deberían ser sumados. Pero hay un número infinito de ellos y la adición de números sólo está definida para un número finito de términos.

Notemos por S_n la suma de los n primeros términos de la serie; la llamaremos suma *parcial* n -ésima. Como resultado obtenemos una sucesión de números

$$\begin{aligned} S_1 &= u_0, \\ S_2 &= u_0 + u_1, \\ &\dots\dots\dots, \\ S_n &= u_0 + u_1 + u_2 + \dots + u_{n-1}, \\ &\dots\dots\dots, \end{aligned}$$

y se puede hablar de una magnitud variable S_n , donde $n = 1, 2, \dots$

Se dice que la serie es *convergente* si, para $n \rightarrow \infty$, la variable S_n tiende a un límite finito

$$\lim_{n \rightarrow \infty} S_n = S.$$

Este límite se llama *suma de la serie*, y en este caso se escribe

$$S = u_0 + u_1 + u_2 + \dots.$$

Pero si cuando $n \rightarrow \infty$ no existe límite de S_n , entonces la serie se dice que es *divergente*, y en este caso no tiene sentido hablar de su suma³⁷. Sin embargo, si todos los u_n tienen el mismo signo, se acostumbra a decir que la suma de la serie es igual a infinito, añadiendo el signo correspondiente.

Como ejemplo, consideremos la serie

$$1 + x + x^2 + \dots,$$

³⁷ Observemos que también es posible dar definiciones generalizadas de la suma de una serie, en virtud de las cuales es posible asignar a ciertas series divergentes un concepto más o menos natural de «suma generalizada». Tales series se dicen sumables. Las operaciones con sumas generalizadas de series divergentes son a veces muy útiles.

cuyos términos forman una progresión geométrica de razón x . La suma de los n primeros términos es igual a

$$S_n(x) = \frac{1 - x^n}{1 - x} \quad (x \neq 1); \quad (55)$$

si $|x| < 1$ esta suma tiene el límite siguiente:

$$\lim_{n \rightarrow \infty} S_n(x) = \frac{1}{1 - x},$$

y así para $|x| < 1$ podemos escribir

$$\frac{1}{1 - x} = 1 + x + x^2 + \cdots$$

Si $|x| > 1$, entonces evidentemente

$$\lim_{n \rightarrow \infty} S_n(x) = \infty,$$

y la serie diverge. La misma situación se tiene también para $x = 1$, como se comprueba inmediatamente sin utilizar la fórmula (55), que carece de sentido para $x = 1$. Finalmente, si $x = -1$ las sumas parciales toman los valores $+1$ y 0 alternativamente, por lo que esta serie es también divergente.

A cada serie corresponde una sucesión dada de sumas parciales $S_1, S_2, S_3, \dots$, y la convergencia de la serie depende de que esta sucesión de sumas tenga límite. Recíprocamente, a una sucesión arbitraria de números $S_1, S_2, S_3, \dots$ corresponde una serie

$$S_1 + (S_2 - S_1) + (S_3 - S_2) + \cdots,$$

cuyas sumas parciales serán los términos de la sucesión. Así, la teoría de las sucesiones puede reducirse a la de las series, y recíprocamente, aunque cada una de estas teorías tiene un significado propio. En algunos casos conviene estudiar la sucesión directamente, y en otros considerar la serie correspondiente.

Las series sirvieron durante mucho tiempo como método importante para representar distintas entidades (sobre todo funciones) y calcular sus valores. Naturalmente, el criterio de los matemáticos

en lo referente a las series ha cambiado con el paso del tiempo, de acuerdo con los cambios que ha ido experimentando la teoría de los infinitésimos. La definición anterior de convergencia y divergencia de una serie fue formulada a comienzos del siglo pasado al mismo tiempo que el concepto de límite.

Si la serie converge, su término general tiende a cero cuando n tiende a infinito, puesto que

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} (S_{n+1} - S_n) = S - S = 0.$$

Los ejemplos que se analizarán en las secciones siguientes, demuestran que la afirmación recíproca es, en general, falsa. Pero el criterio no deja de ser útil, puesto que proporciona una condición *necesaria* para la convergencia de una serie. Por ejemplo, la divergencia de una progresión geométrica de razón $x > 1$ se sigue inmediatamente del hecho de que su término general no tiende a cero.

Si la serie es de términos positivos, su suma parcial S_n aumenta con n , y sólo pueden ocurrir dos casos: o bien la variable S_n se hace mayor que cualquier número dado A a partir de un n suficientemente grande, y en ese caso $\lim_{n \rightarrow \infty} S_n = \infty$, por lo que la serie diverge; o bien existe un número A tal que para todo n el valor de S_n no supera a A ; pero entonces la variable S_n tiende necesariamente a un límite finito dado, no mayor que A , y la serie es convergente.

Convergencia de una serie

La cuestión de si una serie dada converge o diverge puede con frecuencia resolverse por comparación con otra serie. El criterio que se suele utilizar es el siguiente:

Dadas dos series de términos positivos

$$\begin{aligned} u_0 + u_1 + u_2 + \cdots, \\ v_0 + v_1 + v_2 + \cdots, \end{aligned}$$

tales que para todos los valores de n , desde uno de ellos en adelante, se verifica la desigualdad

$$u_n \leq v_n,$$

entonces la convergencia de la segunda serie implica la convergencia de la primera, y la divergencia de la primera implica la divergencia de la segunda.

Por ejemplo, consideremos la llamada serie armónica:

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \cdots.$$

Sus términos son no menores que los términos correspondientes de la serie

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \underbrace{\frac{1}{16} + \cdots + \frac{1}{16}}_{8 \text{ veces}} + \cdots,$$

en la cual la suma de los términos subrayados es igual a $\frac{1}{2}$.

Es claro que la suma S_n de la segunda serie tiende a infinito, y por consiguiente la serie armónica diverge.

La serie

$$1 + \frac{1}{2^a} + \frac{1}{3^a} + \frac{1}{4^a} + \cdots, \quad (56)$$

en la que a es un número positivo menor que la unidad, también diverge, puesto que para todo n se cumple

$$\frac{1}{n^a} \geq \frac{1}{n} \quad (0 < a < 1).$$

Por otra parte, se puede demostrar que para $a > 1$ la serie (56) es convergente. Lo demostraremos sólo para el caso $a \geq 2$; para ello consideremos la serie de términos positivos

$$\left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) + \cdots.$$

Su suma es uno, puesto que las sumas parciales S_n son iguales a

$$S_n = 1 - \frac{1}{n+1} \rightarrow 1 \quad (n \rightarrow \infty).$$

Por otra parte, el término general de esta serie satisface la desigualdad

$$\frac{1}{n-1} - \frac{1}{n} = \frac{1}{(n-1)n} > \frac{1}{n^2},$$

de lo cual se deduce que la serie

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots$$

converge. Así, la serie (56) también convergerá para $\alpha > 2$.

Daremos sin demostración otro útil criterio de convergencia y divergencia de series de términos positivos, llamado criterio de d'Alembert.

Supongamos que cuando n tiende a infinito, el cociente $(u_{n+1})/u_n$ tiene límite q . Entonces para $q < 1$ la serie es convergente, mientras que para $q > 1$ es divergente. Si $q = 1$, nada puede decirse acerca de la convergencia.

La suma de un número finito de sumandos no varía al permutar éstos, lo cual no siempre es cierto para series infinitas. Existen series convergentes cuya suma varía al permutar sus términos, llegando incluso a convertirse en series divergentes. Las series con sumas inestables de esta clase no poseen una de las propiedades fundamentales de las sumas ordinarias: la permutabilidad de los sumandos. Importa, pues, distinguir las series que conservan esta propiedad. Se las conoce con el nombre de series absolutamente convergentes. La serie

$$u_0 + u_1 + u_2 + u_3 + \cdots$$

se dice *absolutamente convergente* si la serie

$$|u_0| + |u_1| + |u_2| + |u_3| + \cdots$$

de los valores absolutos de sus términos es convergente. Se puede demostrar que una serie absolutamente convergente es siempre convergente; en otras palabras, que sus sumas parciales S_n tienden a un límite finito. Es evidente que toda serie convergente cuyos términos tengan todos el mismo signo, es absolutamente convergente.

La serie

$$\frac{\text{sen } x}{1^2} + \frac{\text{sen } 2x}{2^2} + \frac{\text{sen } 3x}{3^2} + \cdots$$

es un ejemplo de serie absolutamente convergente, puesto que los términos de la serie

$$\left| \frac{\text{sen } x}{1^2} \right| + \left| \frac{\text{sen } 2x}{2^2} \right| + \left| \frac{\text{sen } 3x}{3^2} \right| + \cdots$$

son no mayores que los correspondientes de la serie convergente

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots$$

Un ejemplo de serie convergente, pero no absolutamente convergente, es el siguiente:

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots,$$

como el lector puede comprobar por sí mismo.

Series de funciones; series uniformemente convergentes

En análisis se manejan con frecuencia series cuyos términos son funciones de x . En las secciones precedentes vimos ya ejemplos de esta clase, por ejemplo, la serie $1 + x + x^2 + x^3 + \cdots$. Para algunos valores de x esta serie converge, pero, para otros, diverge. Particularmente importantes en las aplicaciones son las series de funciones que convergen para todos los valores de x pertenecientes a un cierto intervalo, que puede ser todo el eje real, o la mitad positiva de él, etcétera. Por otra parte, en ocasiones surge la necesidad de derivar tales series término a término, integrarlas, decidir si su suma es continua, etc. Para el caso elemental de la suma de un número finito de términos hay reglas generales muy sencillas. Sabemos que la derivada de una suma de funciones derivables es igual a la suma de sus derivadas; que la integral de una suma de funciones continuas es la suma de sus integrales, y que una suma de funciones continuas es una función continua: todo esto es cierto para la suma de un número finito de términos.

Pero para las series infinitas estas sencillas reglas no son en general válidas. Se pueden dar muchos ejemplos de series de funciones convergentes para las cuales las reglas de integración y derivación término a término no se cumplen. Del mismo modo, la suma de una serie de funciones continuas puede ser discontinua. Sin embargo, muchas series infinitas se comportan igual que las sumas finitas en lo referente a estas reglas.

Profundas investigaciones sobre esta cuestión han demostrado que las reglas anteriores sí son válidas cuando la serie infinita en cuestión no sólo es convergente en cada punto del intervalo de definición (el dominio en el que varía x), sino también *uniformemente*

convergente en todo el intervalo. De este modo cristalizó en el análisis matemático, a mediados del siglo XIX, el importante concepto de la convergencia uniforme de una serie.

Consideremos la serie

$$S(x) = u_0(x) + u_1(x) + u_2(x) + \cdots,$$

cuyos términos son funciones definidas en el intervalo $[a, b]$. Supongamos que para cada valor de x de dicho intervalo la serie converge hacia una cierta suma $S(x)$. La suma de los n primeros términos de la serie

$$S_n(x) = u_0(x) + u_1(x) + \cdots + u_{n-1}(x)$$

es también una cierta función de x definida en $[a, b]$.

Introduzcamos ahora una magnitud η_n , que es igual a la mínima cota superior de los valores ³⁸ $|S(x) - S_n(x)|$, cuando x varía en el intervalo $[a, b]$. Esta magnitud se escribe como sigue ³⁹:

$$\eta_n = \sup_{a \leq x \leq b} |S_n(x) - S(x)|.$$

En el caso de que $S(x) - S_n(x)$ alcance su valor máximo (lo que ocurre, por ejemplo, cuando $S(x)$ y $S_n(x)$ son continuas), η_n es sencillamente el máximo de $|S(x) - S_n(x)|$ en $[a, b]$.

De la supuesta convergencia de la serie, se tiene que para cada valor de x del intervalo $[a, b]$.

$$\lim_{n \rightarrow \infty} |S(x) - S_n(x)| = 0.$$

Pero η_n puede tender o no a cero. Si η_n tiende a cero cuando $n \rightarrow \infty$, la serie se dice que es *uniformemente convergente*; y en el caso opuesto, no uniformemente convergente. En el mismo sentido se puede hablar de la convergencia uniforme o no uniforme de una sucesión de funciones $S_n(x)$ sin interpretarlas necesariamente como sumas parciales de una serie.

³⁸ Ver capítulo 14.

³⁹ Sup es una abreviatura de la palabra latina *supremum*.

Ejemplo 1

La serie de funciones

$$\frac{1}{x+1} - \frac{1}{(x+1)(x+2)} - \frac{1}{(x+2)(x+3)} - \cdots,$$

que se considera definida sólo para valores no negativos de x , es decir sobre la semirecta $[0, \infty)$, puede escribirse en la forma

$$\frac{1}{x+1} + \left(\frac{1}{x+2} - \frac{1}{x+1} \right) + \left(\frac{1}{x+3} - \frac{1}{x+2} \right) + \cdots,$$

donde se ve que sus sumas parciales son iguales a

$$S_n(x) = \frac{1}{x+n}$$

y que

$$\lim_{n \rightarrow \infty} S_n(x) = 0.$$

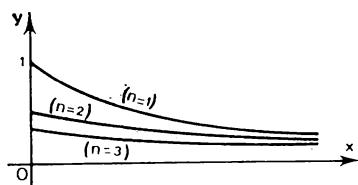


FIG. 37

Así, pues, la serie es convergente para todo x no negativo y tiene como suma $S(x) = 0$. Además,

$$\eta_n = \sup_{0 \leq x < \infty} |S_n(x) - S(x)| = \sup_{0 \leq x < \infty} \frac{1}{x+n} = \frac{1}{n} \rightarrow 0 \quad (n \rightarrow \infty),$$

por lo que la serie converge uniformemente a cero sobre el semieje $[0, \infty)$. La figura 37 muestra las gráficas de algunas de sus sumas parciales $S_n(x)$.

Ejemplo 2

La serie

$$x + x(x-1) + x^2(x-1) + \cdots$$

puede escribirse en la forma

$$x + (x^2 - x) + (x^3 - x^2) + \cdots,$$

de donde

$$S_n(x) = x^n,$$

y por tanto

$$\lim_{n \rightarrow \infty} S_n(x) = \begin{cases} 0, & \text{si } 0 \leq x < 1 \\ 1, & \text{si } x = 1. \end{cases}$$

La suma de la serie es discontinua en el intervalo $[0, 1]$, con una discontinuidad en el punto $x = 1$. La diferencia $|S_n(x) - S(x)|$ es menor que 1 para todo x en $[0, 1]$, pero puede tomar valores tan próximos a la unidad como queramos, con tal de elegir x suficientemente próximo al valor $x = 1$. Así,

$$\eta_n = \sup_{0 \leq x < 1} |S_n(x) - S(x)| = 1$$

para todo $n = 1, 2, \dots$. La serie converge no uniformemente en el intervalo $[0, 1]$. La figura 38 muestra algunas de las gráficas de la función $S_n(x)$. La gráfica de la suma de la serie consta del segmento $0 \leq x < 1$ del eje x , exceptuando el punto del extremo de la derecha, y del punto $(1, 1)$.

Este ejemplo muestra que la suma de una serie no uniformemente convergente de funciones continuas puede ser, de hecho, una función discontinua.

Por otra parte, si consideramos la serie en el intervalo $0 \leq x \leq q$, con $q < 1$, entonces

$$\eta_n = \sup_{0 \leq x \leq q} |S_n(x) - S(x)| = \max_{0 \leq x \leq q} x^n = q^n \xrightarrow{n \rightarrow \infty} 0,$$

por lo que en este intervalo la serie converge uniformemente y su suma es continua. El hecho de que la suma de una serie uniforme-

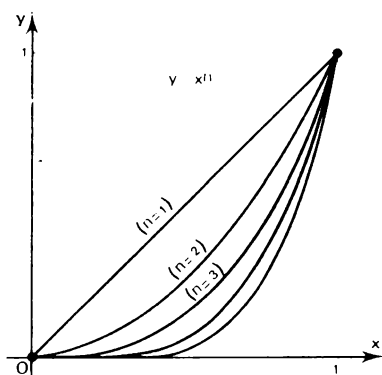


FIG. 38

mente convergente de funciones continuas sea también una función continua es de validez general, como ya se indicó, y puede demostrarse rigurosamente.

Ejemplo 3

La suma de los n primeros términos de la serie $S_n(x)$ tiene la gráfica representada por la línea quebrada de la figura 39. Evidentemente, para todo n tenemos $S_n(0) = 0$; pero si $0 < x \leq 1$, entonces para todo $n \geq 1/x$ tendremos $S_n(x) = 0$, y por consiguiente para x en el intervalo $[0, 1]$ será:

$$S(x) = \lim_{n \rightarrow \infty} S_n(x) = 0.$$

Por otra parte,

$$\begin{aligned} \eta_n &= \sup_{0 \leq x \leq 1} |S_n(x) - S(x)| = \\ &= \sup |S_n(x)| = n^2 \end{aligned}$$

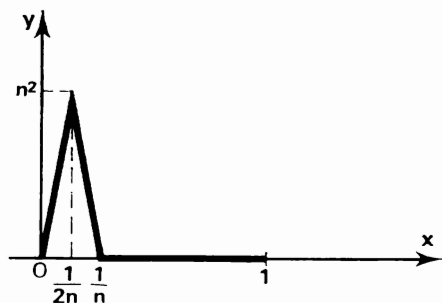


FIG. 39

Así, η_n no sólo no tiende a cero, sino que incluso tiende a infinito. Vemos ahora que la serie correspondiente a esta sucesión $S_n(x)$ no puede ser integrada término a término en el intervalo $[0, 1]$, puesto que

$$\int_0^1 S(x) dx = 0, \quad \int_0^1 S_n(x) dx = \frac{1}{2} n^2 \frac{1}{n} = \frac{n}{2},$$

por lo que la serie

$$\int_0^1 S_1(x) dx + \int_0^1 [S_2(x) - S_1(x)] dx + \int_0^1 [S_3(x) - S_2(x)] dx + \cdots$$

se reduce a la serie divergente

$$\frac{1}{2} + \left(\frac{2}{2} - \frac{1}{2} \right) + \left(\frac{3}{2} - \frac{2}{2} \right) + \left(\frac{4}{2} - \frac{3}{2} \right) + \cdots$$

A continuación damos sin demostración las propiedades fundamentales de las series uniformemente convergentes:

1. La suma de una serie de funciones continuas que es uniformemente convergente en el intervalo $[a, b]$ es una función continua en este intervalo.

2. Si la serie de funciones continuas

$$S(x) = u_0(x) + u_1(x) + u_2(x) + \cdots \quad (57)$$

converge uniformemente en el intervalo $[a, b]$, puede ser integrada término a término en este intervalo; esto es, para todo x_1, x_2 en $[a, b]$ tenemos la igualdad

$$\int_{x_1}^{x_2} S(t) dt = \int_{x_1}^{x_2} u_0(t) dt + \int_{x_1}^{x_2} u_1(t) dt + \cdots.$$

3. Si en el intervalo $[a, b]$ la serie (57) converge y las funciones $u_k(x)$ tienen derivada continua, la igualdad

$$S'(x) = u'_0(x) + u'_1(x) + u'_2(x) + \cdots, \quad (58)$$

obtenida por derivación término a término de (57), será válida en el intervalo $[a, b]$ si la serie del segundo miembro en (58) converge uniformemente.

Series de potencias

En el §9 se dijo que una función $f(x)$, definida en un intervalo $[a, b]$, se llama analítica si en este intervalo tiene derivadas de cualquier orden y si en un entorno suficientemente pequeño de cualquier punto x_0 del intervalo $[a, b]$ puede ser desarrollada en serie de Taylor convergente

$$f(x) = f(x_0) + \frac{f'(x_0)}{1} (x - x_0) + \frac{f''(x_0)}{2!} (x - x_0)^2 + \cdots. \quad (59)$$

Si introducimos la notación

$$a_n = \frac{f^{(n)}(x_0)}{n!},$$

esta serie puede escribirse en la forma siguiente:

$$f(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots. \quad (60)$$

Una serie de esta clase, en la que los números $a_1, a_2, \dots$ son constantes independientes de x , se llama *serie de potencias*.

Como ejemplo consideremos la serie de potencias

$$1 + x + x^2 + x^3 + \dots, \quad (61)$$

cuyos términos constituyen una progresión geométrica.

Sabemos que para todos los valores de x del intervalo $-1 < x < 1$ esta serie converge, y su suma es igual a

$$S(x) = \frac{1}{1-x}.$$

Para los restantes valores de x la serie diverge.

Se ve fácilmente también que la diferencia entre la suma de la serie y la suma de sus n primeros términos viene dada por la fórmula

$$S(x) - S_n(x) = \frac{x^{n+1}}{1-x}, \quad (62)$$

y que si $-q \leq x \leq q$, donde q es un número positivo menor que la unidad, entonces

$$\eta_n = \max |S(x) - S_n(x)| = \frac{q^{n+1}}{1-q}.$$

De aquí se deduce claramente que η_n tiende a cero al aumentar n indefinidamente, por lo que la serie es uniformemente convergente en el intervalo $-q \leq x \leq q$ para todos los valores de $q < 1$.

Se comprueba fácilmente que la función

$$S(x) = \frac{1}{1-x}$$

tiene derivada de orden n , y que ésta es igual a

$$S^{(n)}(x) = \frac{n!}{(1-x)^{n+1}},$$

de donde

$$S^{(n)}(0) = n!,$$

y la suma de los n primeros términos de la serie de Taylor de la función $S(x)$ coincide exactamente, para $x_0 = 0$, con la suma de los n primeros términos de la serie (59). Además, sabemos que el resto o término residual de la fórmula, dado por la igualdad (62), tiende a cero cuando n crece indefinidamente, para todo x del intervalo $-1 < x < 1$. Así, pues, hemos demostrado que la serie (61) es la serie de Taylor de su suma $S(x)$.

Veamos otro hecho interesante. En el intervalo de convergencia $-1 < x < 1$ de la serie, elijamos un punto arbitrario x_0 . Se comprueba fácilmente que para todo x suficientemente próximo a x_0 , por ejemplo, para todo x que satisfaga la desigualdad

$$\frac{|x - x_0|}{1 - x_0} < 1,$$

tenemos la igualdad

$$\begin{aligned} S(x) &= \frac{1}{1-x} = \frac{1}{1-x_0} \left(1 - \frac{x-x_0}{1-x_0} \right)^{-1} = \\ &= \frac{1}{1-x_0} \left[1 + \frac{x-x_0}{1-x_0} + \left(\frac{x-x_0}{1-x_0} \right)^2 + \cdots \right] = \\ &= \frac{1}{1-x_0} + \frac{x-x_0}{(1-x_0)^2} + \frac{(x-x_0)^2}{(1-x_0)^3} + \cdots \end{aligned} \quad (63)$$

El lector puede comprobar sin dificultad que

$$\frac{S^{(n)}(x_0)}{n!} = \frac{1}{(1-x_0)^{n+1}}.$$

Por tanto, la serie (63) es la serie de Taylor de su suma $S(x)$ y converge hacia ella en un entorno suficientemente pequeño de cualquier punto x_0 que pertenezca al intervalo de convergencia de (61). Puesto que el punto x_0 es arbitrario, esto significa que la función $S(x)$ es analítica en el intervalo.

Todos estos hechos que hemos analizado para una serie de potencias particular, la serie (61), se cumplen para todas las series de potencias ⁴⁰. Es decir, para toda serie de potencias de la forma (60),

⁴⁰ Para una información más detallada sobre este punto consúltese el capítulo 9.

en la que las constantes a_k se eligen mediante una ley determinada, existe un cierto número no negativo R (que puede también ser infinito), llamado *radio de convergencia de la serie* (60), con las propiedades siguientes:

1. Para todos los valores de x del intervalo $x_0 - R < x < x_0 + R$, que se llama intervalo de convergencia, la serie converge y su suma $S(x)$ es una función analítica de x en el intervalo. La convergencia es uniforme para todo intervalo $[a, b]$ contenido completamente en el intervalo de convergencia. La serie en sí es la serie de Taylor de su suma.

2. En los puntos extremos del intervalo de convergencia, la serie puede ser convergente o divergente, según cada caso particular. Pero siempre será divergente fuera del intervalo cerrado $x_0 - R \leq x \leq x_0 + R$.

Sugerimos al lector que considere las series de potencias

$$1 + \frac{x}{1} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

$$1 + x + 2!x^2 + 3!x^3 + \cdots,$$

$$1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots$$

y compruebe que sus radios de convergencia son respectivamente infinito, cero y uno.

De la definición dada anteriormente se deduce que toda función analítica puede ser desarrollada, en un entorno suficientemente pequeño de uno de los puntos en donde está definida, en una serie de potencias que converge hacia la función. Recíprocamente, de lo dicho se deduce que la suma de toda serie de potencias cuyo radio de convergencia no es cero, es una función analítica en su intervalo de convergencia.

Vemos, pues, que las series de potencias están orgánicamente relacionadas con las funciones analíticas. Podríamos incluso decir que en su intervalo de convergencia las series de potencias son el modo natural de representar las funciones analíticas, y también, por tanto, el medio natural de aproximar las funciones analíticas mediante polinomios algebraicos ⁴¹.

⁴¹ Las aproximaciones por encima de los límites del intervalo de convergencia de una serie de potencias requieren otros métodos (ver capítulo 12).

Por ejemplo, del hecho de que la función $1/(1-x)$ puede desarrollarse en serie de potencias

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots,$$

que es convergente en el intervalo $-1 < x < 1$, se deduce que la serie de potencias es uniformemente convergente en un intervalo arbitrario $-a \leq x \leq a$ con $a < 1$, y esto implica la posibilidad de aproximar la función en el intervalo $[-a, a]$ por las sumas parciales de la serie con cualquier grado de aproximación deseado.

Supongamos que se nos pide aproximar la función $1/(1-x)$ por polinomios en el intervalo $[-1/2, 1/2]$ con una aproximación de 0,01. Observemos que para todo x de este intervalo se tiene la desigualdad

$$\left| \frac{1}{1-x} - 1 - x - \cdots - x^n \right| = |x^{n+1} + x^{n+2} + \cdots| \leq \\ \leq |x|^{n+1} + |x|^{n+2} + \cdots \leq \frac{1}{2^{n+1}} + \frac{1}{2^{n+2}} + \cdots = \frac{1}{2^n},$$

y puesto que $2^6 = 64$, y $2^7 = 128$, el polinomio deseado, que aproxima la función en todo el intervalo $[-1/2, 1/2]$ con una precisión de 0,01, tendrá la forma

$$\frac{1}{1-x} \approx 1 + x + x^2 + \cdots + x^7.$$

He aquí otra propiedad extremadamente útil de las series de potencias: pueden ser derivadas término a término en todo punto del intervalo de convergencia. Esta propiedad tiene muchas aplicaciones en la resolución de numerosos problemas de la matemática.

Por ejemplo, se pide encontrar la solución de la ecuación diferencial $y' = y$ con la condición inicial $y(0) = 1$. Buscaremos la solución en forma de serie de potencias,

$$y = a_0 + a_1x + a_2x^2 + \cdots.$$

Debido a la condición inicial, es preciso poner $a_0 = 1$. Suponiendo que esta serie converja, se puede derivar término a término; como resultado se tiene

$$y' = a_1 + 2a_2x + 3a_3x^2 + \cdots.$$

Si sustituimos estas dos series en la ecuación diferencial e igualamos entre sí los coeficientes de las potencias iguales de x , obtenemos

$$a_k = \frac{1}{k!} \quad (k = 1, 2, \dots)$$

y la solución buscada tiene la forma

$$y = 1 + \frac{x}{1} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Como se sabe, esta serie converge para todos los valores de x y su suma es igual a $y = e^x$.

En este caso se ha obtenido una serie cuya suma es una función elemental conocida. Pero no siempre sucede esto; puede ocurrir que la serie de potencias convergente obtenida tenga una suma que no sea una función elemental. Un ejemplo es la serie

$$y_p(x) = x^p \left[1 - \frac{x^2}{2(2p+2)} + \frac{x^4}{2 \cdot 4(2p+2)(2p+4)} - \dots \right],$$

que se obtiene como solución de la ecuación diferencial de Bessel, de gran importancia en las aplicaciones prácticas. De este modo las series de potencias pueden servir para definir funciones.

Parte 2

Capítulo 3

GEOMETRIA ANALITICA

§1. *Introducción*

En la primera mitad del siglo xvii nació una rama completamente nueva de la matemática, la geometría analítica, que vino a establecer un nexo entre las curvas del plano y las ecuaciones algebraicas con dos incógnitas.

Fue un hecho bastante raro éste que tuvo lugar en la matemática: en cuestión de una o dos décadas apareció una rama completamente nueva de la matemática, basada en una idea muy sencilla que hasta entonces no había recibido la atención necesaria. La aparición de la geometría analítica en la primera mitad del siglo xvii no fue accidental. La transición en Europa a los nuevos métodos capitalistas de producción requirió el progreso de casi todas las ciencias. Poco tiempo antes, Galileo y otros científicos habían comenzado a elaborar la mecánica contemporánea; en todas las regiones de las ciencias naturales se habían ido acumulando datos empíricos y perfeccionando los medios de observación, y en lugar de las teorías escolásticas, ya anticuadas, se habían creado otras nuevas. En astronomía los principales científicos habían admitido por fin las enseñanzas de Copérnico. El rápido desarrollo de la navegación de altura necesitaba urgentemente de conocimientos más avanzados de astronomía y de mecánica.

El arte de la guerra necesitaba también de la mecánica. Las elipses y parábolas, cuyas propiedades geométricas como secciones cónicas conocían ya los griegos con perfecto detalle desde hacía casi dos mil años, dejaron de ser propiedad exclusiva de la geometría, como ocurría entre aquéllos. Después que Kepler descubriera que los planetas giran alrededor del Sol en elipses, y Galileo que una piedra lanzada al aire describe una parábola, fue necesario calcular estas elipses y determinar la parábola que recorre una bala disparada por un cañón; fue necesario descubrir la ley según la cual la presión atmosférica, descubierta por Pascal, decrece con la altura; preciso fue también calcular el volumen de diversos cuerpos, etc.

Todas estas cuestiones hicieron surgir casi simultáneamente tres ciencias matemáticas enteramente nuevas: la geometría analítica, el cálculo diferencial y el cálculo integral, incluida la resolución de las ecuaciones diferenciales más sencillas.

Estos tres nuevos campos cambiaron cualitativamente la faz de la matemática, e hicieron posible la resolución de problemas antes inimaginables.

En la primera mitad del siglo XVII, concretamente a comienzos de los años 1600, un grupo constituido por los más relevantes matemáticos empezó a vislumbrar la idea de la geometría analítica, pero fueron dos de ellos, en particular, quienes vieron claramente la posibilidad de crear una nueva rama de la matemática: Pierre Fermat, consejero del parlamento de la ciudad francesa de Toulouse y matemático de fama mundial, y el famoso filósofo francés René Descartes. Descartes está considerado como el principal creador de la geometría analítica. El fue quien, como filósofo, planteó el problema de la generalidad absoluta de esta geometría. Descartes fue el autor del gran tratado filosófico «Discurso del método de dirigir correctamente la razón y buscar la verdad en las ciencias, con aplicaciones a la dióptrica, meteorología y geometría».

La última parte de este trabajo, titulada «Geometría» y publicada en 1637, contiene una exposición bastante completa, aunque algo confusa, de la teoría matemática que desde entonces se ha venido llamando geometría analítica.

§2. *Los dos conceptos fundamentales de Descartes*

Descartes deseaba crear un método que pudiera aplicarse a la resolución de todos los problemas de la geometría, esto es, que suministrara un método general de resolución. La teoría de Descartes se

basa en dos conceptos: el de las coordenadas y el de representar en forma de curva plana cualquier ecuación algebraica con dos incógnitas, valiéndose para ello del método de las coordenadas.

El concepto de las coordenadas

Por *coordenadas de un punto* en el plano entendía Descartes la abscisa y ordenada de este punto, es decir, los valores numéricos x e y de sus distancias (con los signos correspondientes) a dos rectas mutua-

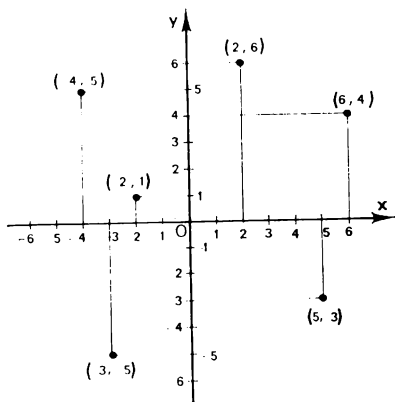


FIG. 1

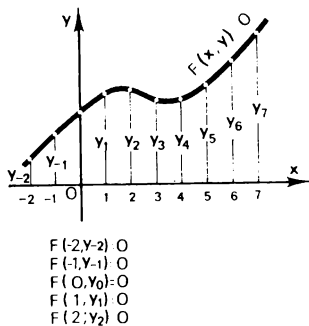


FIG. 2

mente perpendiculares (ejes de coordenadas) elegidas en este plano (ver capítulo 2). El punto de intersección de los ejes de coordenadas, es decir, el punto que tiene por coordenadas $(0, 0)$, se llama *origen*.

Con la introducción de las coordenadas Descartes permitió construir lo que cabría llamar una «aritmétización» del plano. En lugar de determinar un punto geoméricamente, basta con dar un par de números x e y , y recíprocamente (Fig. 1).

Comparación entre las ecuaciones con dos incógnitas y las curvas de un plano

El segundo concepto de Descartes es el siguiente. Antes de Descartes, cuando se planteaba una ecuación algebraica con dos incóg-

nititas $F(x, y) = 0$ se decía que el problema era indeterminado, puesto que resultaba imposible determinar estas incógnitas a partir de la ecuación; podía asignarse cualquier valor a una de ellas, por ejemplo a x , y sustituirlo en la ecuación; el resultado era una ecuación con una sola incógnita y , para la cual, en general, la ecuación tenía solución. Cada x *arbitrariamente elegida*, junto con la y calculada, satisfaría la ecuación dada. Por consiguiente, se estimaba que una tal ecuación «indeterminada» carecía de interés.

Descartes consideró la situación desde un ángulo diferente. Propuso que, en una ecuación con dos incógnitas, la x fuese considerada como la abscisa de un punto, y la correspondiente y como su ordenada. Si se dan entonces distintos valores a la incógnita x , y , para cada valor de x , se calcula la correspondiente y en la ecuación, se obtiene en general un conjunto de puntos que constituyen una curva (Fig. 2)¹.

Así, cada ecuación algebraica con dos variables, $F(x, y) = 0$, va ligada a una curva totalmente determinada del plano, y esta curva representa la totalidad de los puntos del plano cuyas coordenadas satisfacen la ecuación $F(x, y) = 0$.

Esta observación de Descartes *abrió las puertas a una ciencia enteramente nueva*.

Problemas básicos resueltos por la geometría analítica y definición de ésta

La geometría analítica permite: 1) resolver problemas de construcción geométrica mediante el cálculo (por ejemplo, la división de un segmento según una razón dada; ver §3); 2) encontrar la ecuación de una curva definida por una propiedad geométrica (por ejemplo, de una elipse definida por la condición de que la suma de las distancias a dos puntos dados sea constante; ver §7); 3) demostrar nuevos teoremas geométricos algebraicamente (ver, por ejemplo, la deducción de la teoría de Newton de los diámetros, §6); 4) recíprocamente, representar geoméricamente una ecuación algebraica e ilustrar así sus propiedades algebraicas (ver, por ejemplo, la solución de las ecuaciones de tercer y cuarto grado como intersección de una parábola con un círculo, §5).

¹ Algunas veces la ecuación no se satisface por ningún punto (x, y) de coordenadas reales; otras, por uno o varios de tales puntos. En este caso se dice que la curva es imaginaria o que se reduce a puntos (ver §7).

Así, la geometría analítica es aquella parte de la matemática que, aplicando el método de las coordenadas, estudia los objetos geométricos por medios algebraicos.

§3. Problemas elementales

Coordenadas de un punto que divide a un segmento según una razón dada

Dadas las coordenadas (x_1, y_1) y (x_2, y_2) de dos puntos M_1 y M_2 , encontrar las coordenadas (x, y) del punto M que divide al segmento M_1M_2 en la razón m a n (Fig. 3). De la semejanza de los triángulos rayados tenemos:

$$\frac{x - x_1}{x_2 - x} = \frac{m}{n},$$

de donde

$$x = \frac{nx_1 + mx_2}{m + n};$$

por otro lado:

$$\frac{y - y_1}{y_2 - y} = \frac{m}{n},$$

de donde

$$y = \frac{ny_1 + my_2}{m + n}.$$

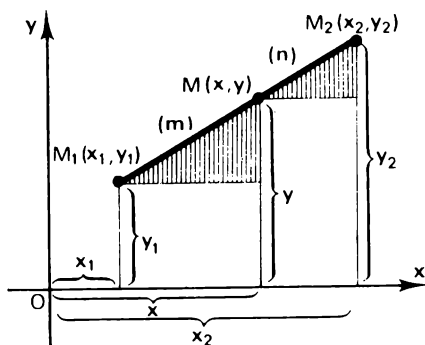


FIG. 3

Distancia entre dos puntos

Hallar la distancia entre dos puntos M_1 y M_2 cuyas coordenadas son (x_1, y_1) y (x_2, y_2) , respectivamente. Del triángulo rayado (Fig. 4) se obtiene, por el teorema de Pitágoras,

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Área de un triángulo

Hallar el área S del triángulo $M_1M_2M_3$ (Fig. 5), siendo las coordenadas de sus vértices (x_1, y_1) , (x_2, y_2) y (x_3, y_3) , respectivamente.

Considerando el área del triángulo como la suma de las áreas de los trapezoides de bases y_1, y_3 e y_3, y_2 , menos el área del trapezoide de base y_1, y_2 , y escribiendo el producto $-(y_1 + y_2)(x_2 - x_1)$ en la forma $(y_1 + y_2)(x_1 - x_2)$, se obtiene

$$S = \frac{1}{2} [(y_1 + y_2)(x_1 - x_2) + (y_2 + y_3)(x_2 - x_3) + (y_3 + y_1)(x_3 - x_1)].$$

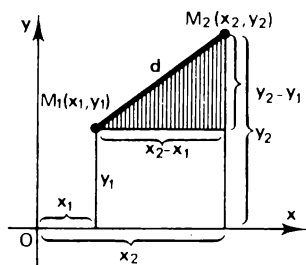


FIG. 4

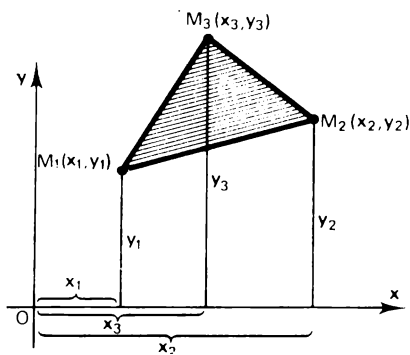


FIG. 5

En estos problemas sólo queda por comprobar que las fórmulas obtenidas siguen siendo válidas en aquellos casos en que una o más coordenadas, o sus diferencias, son negativas. Tales comprobaciones no ofrecen ninguna dificultad.

Determinación de los puntos de intersección de dos curvas

Apoyándonos en la segunda idea fundamental (a saber, que la ecuación $F(x, y) = 0$ representa una curva), resulta particularmente sencillo encontrar los puntos de intersección de dos curvas. Para ello es preciso, evidentemente, resolver simultáneamente las ecuaciones que las representan, pues el par de números x, y obtenidos como solución de estas dos ecuaciones determinará el punto cuyas coordenadas satisfacen ambas ecuaciones, es decir, el punto que pertenece tanto a la primera como a la segunda curva y que es, por consiguiente, el punto de intersección.

La solución de los problemas geométricos con las herramientas de la geometría analítica es, como vemos, muy conveniente por razones prácticas, especialmente porque toda solución se obtiene inmediatamente en forma numérica. *Una geometría de este tipo era exactamente lo que se necesitaba en aquel tiempo.*

§4. Discusión de las curvas representadas por ecuaciones de primer y segundo grado

Ecuaciones de primer grado

Haciendo uso de su segunda idea, Descartes examinó en primer lugar las curvas que corresponden a la ecuación de primer grado,

$$Ax + By + C = 0, \quad (1)$$

es decir, a una ecuación en la que A , B y C son coeficientes numéricos y A y B no son simultáneamente nulos. Comprobó que, en el plano, tal ecuación corresponde siempre a una recta.

Demostraremos que la ecuación (1) siempre representa una recta y, recíprocamente, que toda recta del plano viene representada por una única ecuación de la forma (1). En efecto, supongamos, por ejemplo, que $B \neq 0$; entonces podemos despejar y en la ecuación (1):

$$y = kx + l,$$

donde $k = -\frac{A}{B}$; $l = -\frac{C}{B}$.

Examinemos primero la ecuación $y = kx$. Evidentemente representa una recta que pasa por el origen y forma un ángulo ϕ con el eje x cuya tangente, $\operatorname{tg} \phi$, es k (Fig. 6). En efecto, la ecuación puede escribirse en la forma $y/x = k$, por lo que las coordenadas de todo punto (x, y) de la recta satisfacen la ecuación, mientras que ningún punto (x, y) que no pertenezca a la recta verificará la ecuación, puesto que para un tal punto y/x será mayor o menor que k . Además, si $\operatorname{tg} \phi > 0$, ambas coordenadas x e y son al mismo tiempo positivas o negativas, y si $\operatorname{tg} \phi < 0$, sus signos son opuestos.

Así, la ecuación $y = kx$ representa una recta que pasa por el origen O , y por tanto la ecuación $y = kx + l$ representa también una

recta: la que se obtiene de la anterior por una traslación paralela tal que la ordenada de cada uno de sus puntos se vea incrementada en l (Fig. 7).

Las fórmulas obtenidas anteriormente para determinar las coordenadas de un punto que divide a un segmento en una razón dada, calcular la distancia entre dos puntos dados y hallar el área de un triángulo, junto con lo estudiado sobre la ecuación de una recta, permiten ya resolver un gran número de problemas.

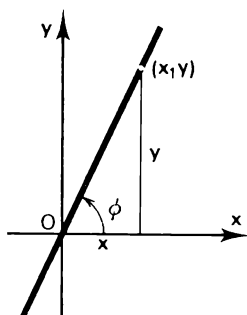


FIG. 6

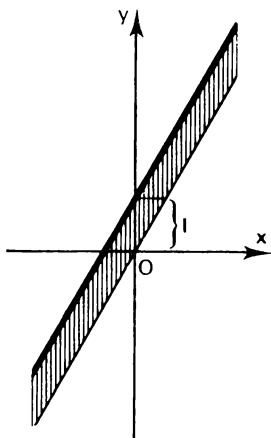


FIG. 7

Ecuaciones de la recta que pasa por uno o dos puntos dados

Sea M_1 un punto de coordenadas x_1, y_1 , y sea k un número dado. La ecuación $y = kx + l$ representa una recta que forma con el eje Ox un ángulo cuya tangente es igual a k y corta al eje Oy a una distancia l de O . Elijamos l de modo que esta recta pase por el punto (x_1, y_1) . Para ello, las coordenadas del punto M_1 deberán satisfacer la ecuación, es decir, se debe verificar $y_1 = kx_1 + l$, y por tanto, $l = y_1 - kx_1$. Sustituyendo l por este valor, se obtiene la ecuación de la recta que pasa por el punto dado (x_1, y_1) y que forma con el eje Ox un ángulo cuya tangente es igual a k (Fig. 8). Esta ecuación es $y = kx + y_1 - kx_1$, o bien

$$y - y_1 = k(x - x_1).$$

Ejemplo

Sea 45° el ángulo formado por la recta y el eje Ox , y sea M el punto de coordenadas $(3, 7)$; la ecuación de la recta correspondiente será (puesto que $\operatorname{tg} 45^\circ = 1$): $y - 7 = 1 \cdot (x - 3)$, o bien: $x - y + 4 = 0$.

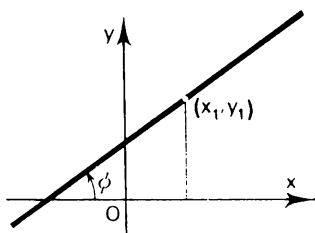


FIG. 8

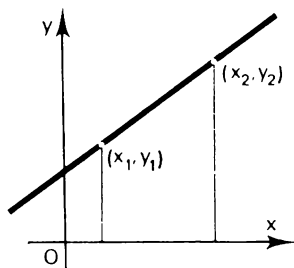


FIG. 9

Si obligamos a que la recta que pasa por el punto (x_1, y_1) pase también por el punto (x_2, y_2) , entonces hay que imponer a k la condición $y_2 - y_1 = k(x_2 - x_1)$. Despejando k y sustituyendo en la ecuación anterior, se obtiene la ecuación de la recta que pasa por dos puntos dados (Fig. 9):

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}.$$

Resultado de Descartes para las ecuaciones de segundo grado

Descartes también estudió la cuestión de qué clase de curvas del plano vienen representadas por las ecuaciones de segundo grado con dos variables, cuya expresión general es

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

y demostró que tal ecuación representa, en general, una elipse, una hipérbola o una parábola; es decir, curvas muy conocidas de los matemáticos de la Antigüedad.

Estas son las aportaciones más importantes de Descartes. Sin embargo, su libro no se limitaba a estos temas; también investigó las ecuaciones de un cierto número de interesantes lugares geométricos, examinó diversos teoremas sobre la transformación de ecuaciones algebraicas, mencionó sin demostración su famosa *ley de los signos* para el número de raíces positivas de una ecuación cuyas raíces son todas reales (ver capítulo 4, §4) y, finalmente, expuso un interesante método para determinar las raíces reales de las ecuaciones de tercer y cuarto grado por intersección de la parábola $y = x^2$ con circunferencias.

§5. *Método de Descartes para la resolución de las ecuaciones algebraicas de tercer y cuarto grado*

Transformación de las ecuaciones de tercer y cuarto grado en ecuaciones de cuarto grado sin término en x^3

Vamos a demostrar que la solución de cualquier ecuación de tercer o cuarto grado puede reducirse a la solución de una ecuación de la forma

$$x^4 + px^2 + qx + r = 0. \quad (2)$$

Sea $z^3 + az^2 + bz + c = 0$ la ecuación general de tercer grado. Sustituyendo $z = x - a/3$, se tiene,

$$(x - a/3)^3 + a(x - a/3)^2 + b(x - a/3) + c = 0.$$

Los términos en x^2 se cancelan al desarrollar los paréntesis, por lo que se obtiene una ecuación de la forma $x^3 + px + q = 0$. Multiplicando esta ecuación por x , queda en la forma (2) con $r = 0$, que admite también la raíz $x_4 = 0$.

Una ecuación de cuarto grado $z^4 + az^3 + bz^2 + cz + d = 0$ puede reducirse a la forma (2) con la sustitución $z = x - a/4$. Por tanto, la solución de todas las ecuaciones de tercer y cuarto grado puede reducirse a la solución de una ecuación de la forma (2).

Solución de las ecuaciones de tercer y cuarto grado por intersección de una circunferencia con la parábola $y = x^2$

Obtengamos en primer lugar la ecuación de una circunferencia de centro (a, b) y radio R . Si (x, y) es uno cualquiera de sus puntos, entonces el cuadrado de su distancia al punto (a, b) es igual a

$(x - a)^2 + (y - b)^2$ (ver §3). Así, la ecuación de la circunferencia en cuestión es

$$(x - a)^2 + (y - b)^2 = R^2.$$

Tratemos ahora de encontrar los puntos de intersección de esta circunferencia con la parábola $y = x^2$. Para ello, en virtud de lo establecido en el §3, es necesario resolver simultáneamente la ecuación de esta circunferencia

$$x^2 + y^2 - 2ax - 2by + a^2 + b^2 - R^2 = 0.$$

y la ecuación de la parábola

$$y = x^2.$$

Sustituyendo la y de la segunda ecuación en la primera obtenemos una ecuación de cuarto grado en x :

$$x^2 + x^4 - 2ax - 2bx^2 + a^2 + b^2 - R^2 = 0$$

o bien

$$x^4 + (1 - 2b)x^2 - 2ax + a^2 + b^2 - R^2 = 0.$$

Si elegimos a , b y R^2 tal que

$$1 - 2b = p, \quad -2a = q, \quad a^2 + b^2 - R^2 = r,$$

obtenemos precisamente una ecuación de la forma (2). Para ello tenemos que tomar

$$a = -\frac{q}{2}, \quad b = \frac{1-p}{2}, \quad R^2 = \frac{q^2}{4} + \frac{(1-p)^2}{4} - r. \quad (3)$$

En la fórmula (3), R^2 puede ser negativo. Sin embargo, en el caso en el que la ecuación (2) tenga una raíz real x_1 , se verificará la siguiente igualdad:

$$x_1^4 + (1 - 2b)x_1^2 - 2ax_1 + a^2 + b^2 - R^2 = 0. \quad (4)$$

Notando x_1^2 por y_1 , la ecuación (4) puede escribirse como

$$x_1^2 + y_1^2 - 2ax_1 - 2by_1 + a^2 + b^2 - R^2 = 0$$

o como

$$(x_1 - a)^2 + (y_1 - b)^2 = R^2.$$

Por tanto, cuando la ecuación (2) tiene una raíz real, el número $R^2 = \frac{(1-p)^2 + q^2}{4} - r$ es *positivo*, la ecuación

$$(x-a)^2 + (y-b)^2 = R^2$$

es la de una circunferencia, y todas las raíces reales de la ecuación (2) son las abscisas de los puntos de intersección de la parábola $y = x^2$ con esta circunferencia. (En el caso $r = 0$, $R^2 = a^2 + b^2$, esta circunferencia pasa por el origen.)

Así, si nos dan los coeficientes p , q , r de la ecuación (2), hay que obtener a , b y R^2 mediante las fórmulas (3); una vez obtenidas, si $R^2 < 0$ la ecuación (2) no tiene raíces reales. Pero si $R^2 \geq 0$, las abscisas de los puntos de intersección de la circunferencia de centro (a, b) y radio R con la parábola $y = x^2$ (dibujada de una vez para todas) dan todas las raíces reales de la ecuación (2); en el caso $R^2 < 0$, la circunferencia resultante no corta a la parábola y la ecuación (2) no tiene raíces reales.

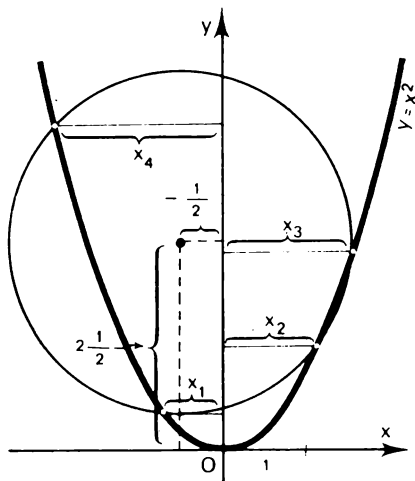


FIG. 10

Ejemplo

Sea la ecuación de cuarto grado:

$$x^4 - 4x^2 + x + \frac{5}{2} = 0.$$

Entonces tenemos

$$a = -\frac{1}{2}, \quad b = \frac{5}{2} = 2\frac{1}{2}; \quad R = \sqrt{\frac{1}{4} + \frac{25}{4} - \frac{5}{2}} = 2.$$

La figura 10 muestra la circunferencia correspondiente y las raíces x_1, x_2, x_3, x_4 de la ecuación dada.

Las secciones anteriores 1.^a, 2.^a, 3.^a y 4.^a contienen, en forma abreviada y algo más moderna, lo esencial del libro de Descartes.

Desde su época hasta ahora la geometría analítica ha experimentado un inmenso desarrollo que ha resultado muy fructífero para muchas partes diferentes de la matemática. En las siguientes secciones de este capítulo intentaremos seguir paso a paso las etapas más importantes de este desarrollo.

En primer lugar, es necesario decir que los inventores del análisis infinitesimal ya conocían el método de Descartes. Tanto si se trataba de tangentes o normales (perpendiculares a las tangentes en el punto de tangencia) a curvas, como de máximos o mínimos de funciones definidas geométricamente, o del radio de curvatura de una curva en un punto dado, etc., se consideraba en primer lugar la ecuación de la curva por el método de Descartes, y luego se obtenían las ecuaciones de la normal, la tangente, etc. Así, el análisis infinitesimal (el cálculo diferencial e integral) hubiera sido inconcebible sin el desarrollo previo de la geometría analítica.

§6. *Teoría general de los diámetros de Newton*

El primer matemático que dio un segundo y nuevo impulso a la geometría analítica fue Newton. En 1704 estudió la teoría de las curvas de tercer orden, es decir, curvas que vienen representadas por ecuaciones algebraicas de tercer grado en dos incógnitas. Al mismo tiempo encontró —entre otros hallazgos— un elegante teorema general sobre los «diámetros» que corresponden a secantes en una dirección dada. Demostró lo siguiente:

Sea una curva de n -ésimo orden, es decir, una curva que viene representada por una ecuación algebraica de n -ésimo grado en dos incógnitas; una recta cualquiera la cortará, en general, en n puntos. Sea M el punto que, perteneciendo a dicha recta secante, sea el «centro de gravedad» de los puntos de su intersección con la curva de n -ésimo orden dada, es decir, el centro de gravedad de un conjunto de n puntos de igual masa que estuvieran situados en dichas intersecciones. Se demuestra que, si se toman todos los posibles conjuntos de secantes paralelas entre sí y para cada una de ellas consideramos estos centros de masa M , entonces para cualquier conjunto dado de secantes paralelas, todos los puntos M se encuentran alineados. Newton llamó a esta recta el «diámetro» de la curva de orden n -ésimo correspondiente a esa dirección concreta de las secantes. Puesto que la demostración de este teorema es bastante fácil con la ayuda de la geometría analítica, la daremos aquí.

Consideremos una curva de orden n y un conjunto de secantes a la curva paralelas entre sí. Elegimos los ejes de coordenadas de modo que estas secantes sean paralelas al eje Ox (Fig. 11). Entonces estas ecuaciones tendrán la forma $y = l$, siendo la constante l diferente para las distintas secantes. Sea $F(x, y) = 0$ la ecuación de la curva de orden n respecto a estos ejes de coordenadas. Es fácil demostrar que, en una transformación de un sistema de coordenadas rectangular

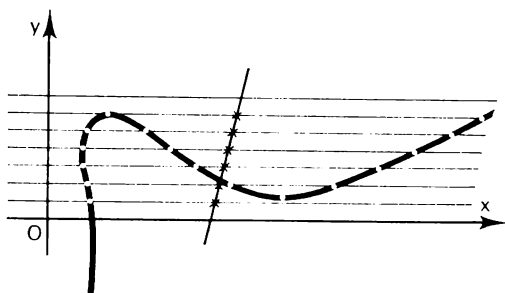


FIG. 11

a otro, la ecuación de la curva varía, pero no así su orden (esto se demostrará en §8). Por tanto, $F(x, y)$ seguirá siendo un polinomio de grado n . Para determinar las abscisas de los puntos de intersección de la curva con la secante $y = l$ es necesario resolver las

ecuaciones simultáneas $F(x, y) = 0$ e $y = l$, obteniéndose como resultado, en general, una ecuación de n -ésimo grado en x :

$$F(x, l) = 0, \quad (5)$$

de la cual obtenemos las abscisas $x_1, x_2, \dots, x_n$. La abscisa x_c del centro de gravedad de los n puntos de intersección es igual, por definición de centro de gravedad, a

$$x_c = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Pero, como se sabe por la teoría de las ecuaciones algebraicas, la suma $x_1 + x_2 + \dots + x_n$ de las raíces de una ecuación es igual al coeficiente de la potencia $(n - 1)$ -ésima de la incógnita x , tomado con signo opuesto y dividido por el coeficiente de la potencia n -ésima de x . Pero, como la suma de los exponentes de x e y en cada uno de los términos de $F(x, y)$ es igual o menor que n , el término x^n no contiene a y , sino que será de la forma Ax^n , donde A es una constante;

y si los términos en x^{n-1} contienen a y , ésta no podrá estar elevada a una potencia superior a la primera; es decir, tendrá la forma $x^{n-1}(By + C)$. Por consiguiente, el coeficiente de x^n es A y el de x^{n-1} es $Bl + C$, y para cualquier l dado tenemos:

$$x_c = - \frac{Bl + C}{nA}.$$

Ahora bien, la secante es paralela al eje Ox , por lo que para todos sus puntos $y = l$, y por consiguiente la ordenada y del centro de gravedad de los puntos de su intersección con la curva de orden n -ésimo dada es también igual a l ; así finalmente obtenemos

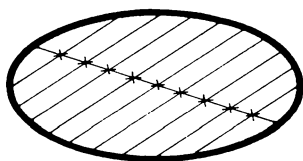


FIG. 12

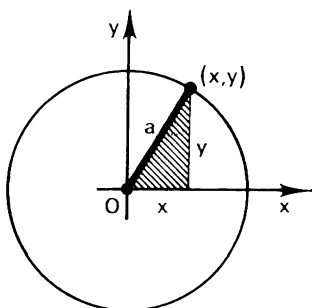


FIG. 13

$nAx_c + By_c + C = 0$; es decir, las coordenadas (x_c, y_c) de los centros de gravedad para todas estas secantes satisfacen una ecuación de primer grado, y por tanto se encuentran sobre una recta.

El caso en que $F(x, y)$ no contenga a x^n admite una demostración análogamente.

Si la curva es de segundo orden ($n = 2$), el centro de gravedad de dos puntos es simplemente su punto medio; de aquí que el lugar geométrico de los puntos de un conjunto de cuerdas paralelas de una curva de segundo orden es una línea recta (Fig. 12), resultado que para la elipse, la hipérbola y la parábola era ya conocido de los antiguos. Pero lo demostraban —aun para estos pocos casos particulares— con razonamientos geométricos sumamente complejos, mien-

tras que ahora, con ayuda de un nuevo teorema desconocido de los antiguos, se demuestra de un modo verdaderamente sencillo.

Tales ejemplos revelan la potencia de la geometría analítica.

§7. *Elipse, hipérbola y parábola*

En esta sección y en las siguientes estudiaremos las curvas de segundo orden. Antes de discutir la ecuación general de segundo grado, es útil examinar alguna de sus formas más sencillas.

Ecuaciones de una circunferencia de centro el origen de coordenadas

En primer lugar, consideraremos la ecuación

$$x^2 + y^2 = a^2.$$

Evidentemente representa una circunferencia de centro el origen y radio a , como se deduce del teorema de Pitágoras aplicado al triángulo rayado de la figura 13, puesto que cualquiera que sea el punto (x, y) que tomemos en esta circunferencia, sus coordenadas x e y satisfarán esta ecuación; y recíprocamente, si las coordenadas x, y de un punto satisfacen la ecuación, entonces el punto pertenece a la circunferencia; es decir, la circunferencia es el conjunto de todos los puntos del plano que satisfacen la ecuación.

La ecuación de una elipse y su propiedad focal

Sean F_1 y F_2 dos puntos dados, y $2c$ la distancia que los separa. Vamos a obtener la ecuación del lugar de todos los puntos M del plano tales que la suma de sus distancias a los puntos F_1 y F_2 es igual a una constante $2a$ (donde, naturalmente, a es mayor que c). Tal curva se llama *elipse*, y los puntos F_1 y F_2 son sus *focos*.

Elijamos un sistema de coordenadas rectangulares tal que los puntos F_1 y F_2 se encuentren sobre el eje Ox y a igual distancia del origen. Entonces las coordenadas de los puntos F_1 y F_2 serán $(c, 0)$ y $(-c, 0)$. Sea M un punto arbitrario de coordenadas (x, y) perteneciente al lugar geométrico en cuestión, y escribamos que la suma de sus distancias a los puntos F_1 y F_2 es igual a $2a$:

$$\sqrt{(x-c)^2 + (y-0)^2} + \sqrt{(x+c)^2 + (y-0)^2} = 2a. \quad (6)$$

Esta ecuación la satisfacen las coordenadas (x, y) de cualquier punto del lugar considerado. Evidentemente, el recíproco es también cierto: cualquier punto cuyas coordenadas satisfacen la ecuación (6) pertenece a este lugar. La ecuación (6) es por tanto la ecuación del lugar. Queda únicamente simplificarla.

Elevando ambos miembros al cuadrado, obtenemos

$$x^2 - 2cx + c^2 + y^2 + 2\sqrt{(x^2 - 2cx + c^2 + y^2)(x^2 + 2cx + c^2 + y^2)} + x^2 + 2cx + c^2 + y^2 = 4a^2,$$

y después de simplificar:

$$x^2 + y^2 + c^2 - 2a^2 = -\sqrt{(x^2 + y^2 + c^2)^2 - 4c^2x^2}.$$

Elevando otra vez al cuadrado, obtenemos

$$(x^2 + y^2 + c^2)^2 - 4a^2(x^2 + y^2 + c^2) + 4a^4 = (x^2 + y^2 + c^2)^2 - 4c^2x^2$$

y simplificando

$$(a^2 - c^2)x^2 + a^2y^2 = (a^2 - c^2)a^2.$$

Hagamos $a^2 - c^2 = b^2$ (lo que es posible puesto que $a > c$); obtenemos así $b^2x^2 + a^2y^2 = a^2b^2$, y dividiendo por a^2b^2 tenemos

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad (7)$$

Las coordenadas (x, y) de cualquier punto M del lugar satisfacen, pues, la ecuación (7).

Por otra parte, puede demostrarse que si las coordenadas de un punto satisfacen la ecuación (7) también satisfacen la ecuación (6). Por tanto, la ecuación (7) es la ecuación de este lugar, es decir, la ecuación de la elipse (Fig. 14).

Este razonamiento es un ejemplo clásico de obtención de la ecuación de una curva cuando ésta viene dada por alguna de sus propiedades geométricas.

El conocido método de trazar una elipse por medio de una cuerda (Fig. 15) se basa en la propiedad de la elipse de que la suma de las distancias de cualquiera de sus puntos a dos puntos dados es constante.

Observación

Para determinar una elipse, en lugar del a propiedad focal considerada, podíamos haber tomado cualquier otra propiedad geométrica característica de ella; por ejemplo, que la elipse es el resultado de una «contracción uniforme» de la circunferencia hacia uno de sus diámetros, o bien cualquier otra propiedad.

Sustituyendo $y = 0$ en la ecuación (7) de la elipse obtenemos $x = \pm a$, es decir: a es la longitud del segmento OA (ver Fig. 14), que se llama *semieje mayor* de la elipse. Análogamente, sustituyendo

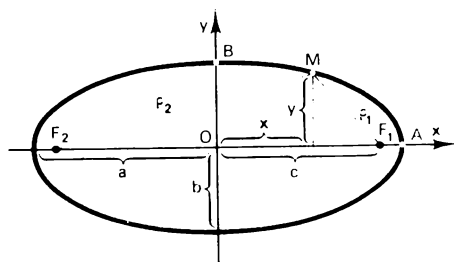


FIG. 14

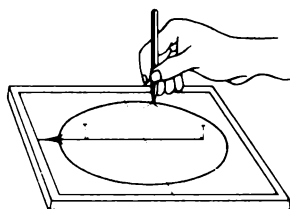


FIG. 15

$x = 0$ obtenemos $y = \pm b$, es decir: b es la longitud del segmento OB , que se llama *semieje menor* de la elipse.

El número $\epsilon = c/a$ se llama *excentricidad* de la elipse, por lo que, al ser $c = \sqrt{a^2 - b^2} < a$, la excentricidad de una elipse será menor que la unidad. En el caso de una circunferencia, $c = 0$, y por tanto $\epsilon = 0$; ambos focos coinciden en un mismo punto, que es el centro de la circunferencia (puesto que $OF_1 = OF_2 = 0$); así, pues, el método anterior de dibujar la curva con un hilo sigue siendo válido.

Ecuaciones del movimiento planetario

Estudiando las observaciones hechas durante mucho tiempo por Tycho Brahe sobre el movimiento del planeta Marte, Kepler descubrió que los planetas giran alrededor del Sol en elipses y de modo tal que el Sol ocupa uno de sus focos (el otro foco permanece vacío

y no juega ningún papel en el movimiento de los planetas alrededor del Sol) (Fig. 16). Kepler observó que el radio focal ϱ de un planeta dado barre siempre la misma área por unidad de tiempo ², y Newton demostró que la existencia de un movimiento de este tipo se deduce matemáticamente de la ley de la inercia (proporcionalidad entre la aceleración y la fuerza) y la ley de la gravitación universal.

Elipse de inercia

Como ejemplo de aplicación de la elipse a un problema técnico, consideremos la llamada elipse de inercia de una placa.

Consideremos una placa de material homogéneo y espesor uniforme, por ejemplo, una placa de cinc de forma arbitraria. La hacemos girar alrededor de un eje de su plano. Un cuerpo en movimiento rectilíneo tiene, como es sabido, una inercia respecto a su movimiento que es proporcional a su masa (independientemente de la forma del cuerpo y de la distribución de la masa). Igualmente un cuerpo que gira alrededor de un eje (por ejemplo, un volante) tiene inercia con respecto a su rotación. Pero en el caso de la rotación la inercia no sólo es proporcional a la masa del cuerpo, sino que también depende de la distribución de la masa del cuerpo con respecto al eje de rotación, puesto que la inercia de la rotación es mayor cuanto más alejada del eje está la masa. Hacer girar rápidamente una barra alrededor de su eje longitudinal (Fig. 17a) no cuesta apenas trabajo, pero si tratamos de hacerla girar alrededor de un eje perpendicular a su longitud, aunque dicho eje pase por su punto medio, comprobamos que, a menos que la barra sea muy ligera, la fuerza que hay que aplicar es considerable (Fig. 17b).

Es posible demostrar que la inercia de un cuerpo respecto a la rotación sobre un eje, lo que llamamos *momento de inercia* del cuerpo respecto al eje, es igual a $\sum r_i^2 m_i$ (donde $\sum r_i^2 m_i$ representa la suma $r_1^2 m_1 + r_2^2 m_2 + \dots + r_n^2 m_n$; para ello hemos descompuesto el cuerpo en elementos muy pequeños, siendo m_i la masa del elemento i -ésimo,

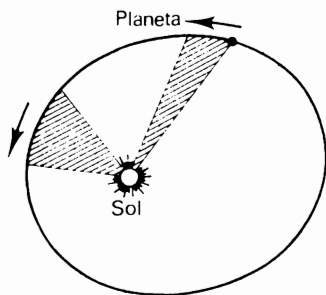


FIG. 16

² Las excentricidades de las órbitas planetarias no son muy grandes, por lo que las órbitas de los planetas son aproximadamente circulares.

y r_i , la distancia de este elemento al eje de rotación, y hemos extendido la suma a todos los elementos del cuerpo).

Volvamos al ejemplo de la placa. Sea O (Fig. 18) un punto de ella. Consideremos el momento de inercia J_u de la placa respecto a un eje de rotación u que pasa por O y que se halla situado en el plano de la placa. Para ello tomamos el punto O como origen de un sistema de coordenadas cartesiano y elegimos ejes arbitrarios Ox y Oy en

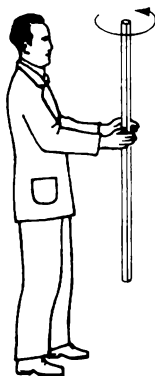


FIG. 17 a

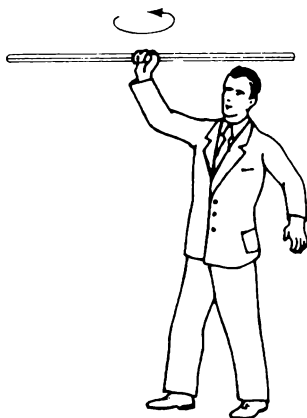


FIG. 17 b

el plano de la placa. Caractericemos el eje de rotación u por el ángulo ϕ que forma con el eje Ox . Es fácil ver (Fig. 19) que

$$r_i = |(x_i \operatorname{tg} \phi - y_i) \cos \phi| = |x_i \operatorname{sen} \phi - y_i \cos \phi|.$$

Por tanto,

$$\begin{aligned} \sum r_i^2 m_i &= \sum (x_i^2 \operatorname{sen}^2 \phi - 2x_i y_i \operatorname{sen} \phi \cos \phi + y_i^2 \cos^2 \phi) m_i = \\ &= \operatorname{sen}^2 \phi \sum x_i^2 m_i - 2 \operatorname{sen} \phi \cos \phi \sum x_i y_i m_i + \cos^2 \phi \sum y_i^2 m_i. \end{aligned}$$

Las cantidades $\operatorname{sen}^2 \phi$, $2 \operatorname{sen} \phi \cos \phi$ y $\cos^2 \phi$ se pueden sacar fuera del signo sumatorio, puesto que son constantes para un eje dado u . Escribamos ahora

$$\sum x_i^2 m_i = A, \quad -\sum x_i y_i m_i = B, \quad \sum y_i^2 m_i = C.$$

Las cantidades A , B y C no dependen de la elección del eje u , sino sólo de la forma de la placa, de la distribución de su masa y de la elección de los ejes de coordenadas Ox y Oy . Por tanto,

$$J_u = A \sin^2 \phi + 2B \sin \phi \cos \phi + C \cos^2 \phi.$$

Considerando todos los posibles ejes u que, pertenecientes al plano de la placa, pasen por el punto O , y tomando sobre cada uno

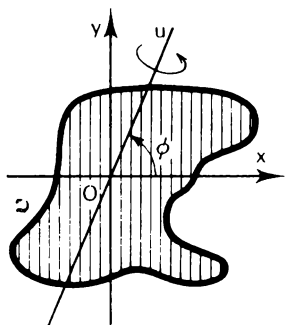


FIG. 18

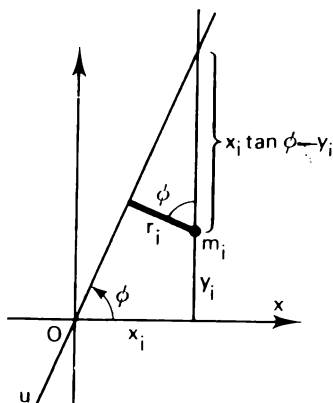


FIG. 19

de estos ejes y a partir del punto O una longitud igual a ϱ , inversa de la raíz cuadrada del momento de inercia J_u de la placa respecto a ese eje, es decir, $\varrho = 1/\sqrt{J_u}$, se obtiene

$$\frac{1}{\varrho^2} = A \sin^2 \phi + 2B \sin \phi \cos \phi + C \cos^2 \phi.$$

Pero

$$x = \varrho \cos \phi, \quad y = \varrho \sin \phi,$$

por lo que la ecuación de este lugar tiene la forma siguiente:

$$Cx^2 + 2Bxy + Ay^2 = 1.$$

El resultado es una curva de segundo orden que, evidentemente, es finita y cerrada; es decir, una elipse (Fig. 20), puesto que todas las otras curvas de segundo orden, como veremos más tarde, o son infinitas o se reducen a un punto.

Hemos llegado así al siguiente resultado importante: cualquiera que sea la forma y tamaño de una placa y la distribución de su masa, el valor de su momento de inercia (con más precisión, el de la cantidad ρ inversamente proporcional a la raíz cuadrada del momento de inercia) con respecto a los distintos ejes del plano de la placa que pasan por el punto dado O , viene caracterizado por una cierta elipse. Esta elipse se llama elipse de inercia de la placa relativa al punto O . Si el punto O es el centro de gravedad de la placa, la elipse se llama entonces elipse central de inercia.

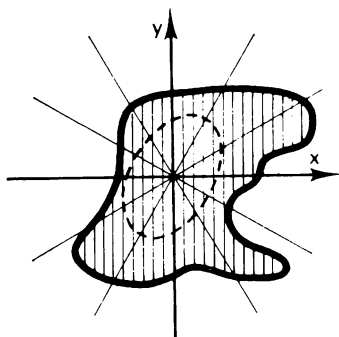


FIG. 20

La elipse de inercia tiene gran importancia en mecánica; en particular, posee aplicaciones importantes en resistencia de mate-

riales. En la teoría de resistencia de materiales se demuestra que la resistencia a la flexión de una viga de sección transversal dada es proporcional al momento de inercia de su sección transversal respecto al eje que, pasando por el centro de gravedad de dicha sección transversal, es perpendicular a la dirección de la fuerza aplicada. Aclaremos esto con un ejemplo. Supongamos que un puente a través de un riachuelo consta de un tablón, y que éste se comba bajo el peso de un peatón que lo atraviesa. Si ahora colocamos ese mismo tablón «de canto», apenas se combará; es decir, un tablón colocado de canto es, por así decirlo, más resistente. Esto se debe a que el momento de inercia de la sección transversal del tablón (esta sección tiene la forma de un rectángulo alargado que podemos imaginar recubierto uniformemente de masa) es mayor respecto al eje perpendicular a su lado mayor que respecto a un eje paralelo a este lado. Si el tablón no se coloca ni de plano ni de canto, sino oblicuamente, o incluso si no tomamos un tablón sino una barra de sección transversal arbitraria —por ejemplo, un raíl— la resistencia a la flexión seguirá siendo proporcional al momento de inercia de su sección transversal

respecto al eje correspondiente. La resistencia de la viga a la flexión viene caracterizada así por la elipse de inercia de su sección transversal.

Para una viga rectangular ordinaria esta elipse tendrá la forma mostrada en la figura 21. La rigidez de un tablón bajo una carga que actúe en la dirección del eje Oz es proporcional a bh^3 .

Las vigas de acero tienen a menudo una sección transversal en

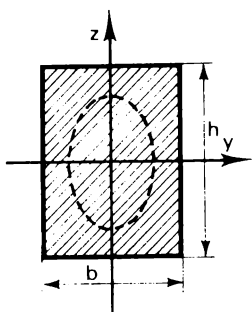


FIG. 21

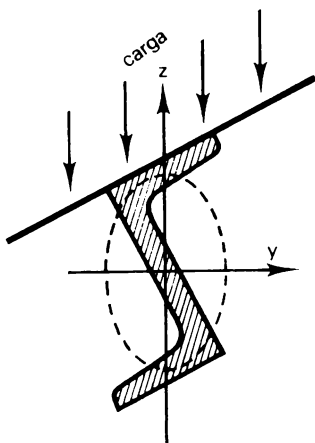


FIG. 22

forma de «zeta»; la sección transversal y la elipse de inercia de este tipo de vigas están representadas en la figura 22. La máxima resistencia a la flexión tiene lugar en la dirección z . Cuando se utilizan, por ejemplo, como vigas de tejado bajo el peso de éste y de la nieve, el trabajo que ejercen contra la flexión ocurre en una dirección que es muy próxima a la más ventajosa.

La hipérbola y su propiedad focal

Consideremos la ecuación

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

que representa una curva llamada *hipérbola*. Si denotamos por c un número tal que $c^2 = a^2 + b^2$, es posible demostrar que una hipérbola es el lugar geométrico de todos los puntos cuya diferencia de

distancia a los puntos F_1 y F_2 (situados sobre el eje Ox y de abscisas c y $-c$) es constante: $\varrho_2 - \varrho_1 = 2a$ (Fig. 23). Los puntos F_1 y F_2 se llaman *focos*.

La parábola y su directriz

Consideremos finalmente la ecuación

$$y^2 = 2px$$

y llamemos *parábola* a la curva correspondiente. El punto F , situado sobre el eje Ox y cuya abscisa es $p/2$, se llama *foco* de la parábola, y la recta $y = -p/2$, paralela al eje Oy , es su *directriz*. Sea M cualquier

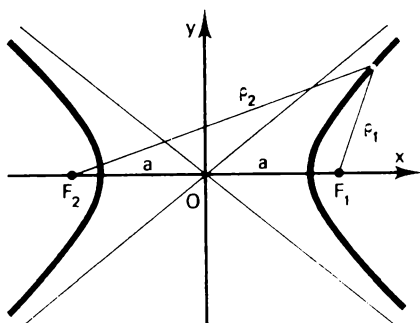


FIG. 23

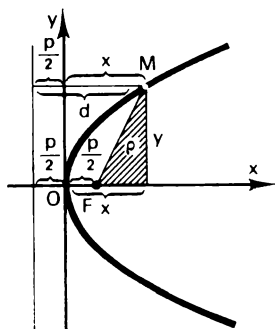


FIG. 24

punto de la parábola (Fig. 24), ϱ la longitud de su radio focal MF , y d la distancia de M a la directriz. Calculemos ϱ y d para el punto M . Del triángulo sombreado obtenemos $\varrho^2 = (x - p/2)^2 + y^2$. Si el punto M se encuentra sobre la parábola, tenemos $y^2 = 2px$; por tanto

$$\varrho^2 = \left(x - \frac{p}{2}\right)^2 + 2px = \left(x + \frac{p}{2}\right)^2.$$

Pero de la figura se desprende directamente que $d = x + p/2$. Por tanto $\varrho^2 = d^2$, es decir, $\varrho = d$. El razonamiento inverso muestra que, si para un punto dado tenemos $\varrho = d$, el punto se encuentra sobre la parábola. Así, la parábola es el lugar geométrico de los puntos que equidistan de un punto dado F (llamado *foco*) y de una recta dada d (llamada *directriz*).

Propiedad de la tangente a una parábola

Examinemos una propiedad importante de la tangente a una parábola y su aplicación a la óptica. Puesto que para una parábola $y^2 = 2px$ tenemos $2y \, dy = 2p \, dx$, se deduce que la derivada, o pendiente de la tangente, es igual a $dy/dx = \operatorname{tg} \phi = p/y$ (Fig. 25).

Por otra parte, de la figura se deduce directamente que

$$\operatorname{tg} \gamma = \frac{y}{x - p/2}.$$

Pero

$$\operatorname{tg} 2\phi = \frac{2p/y}{1 - p^2/y^2} = \frac{2py}{y^2 - p^2} = \frac{2py}{2px - p^2} = \frac{y}{x - p/2},$$

es decir, $\gamma = 2\phi$, y puesto que $\gamma = \phi + \psi$, se deduce $\psi = \phi$. Por tanto, y en virtud de la ley de la reflexión (ángulo de incidencia igual a ángulo de reflexión), un rayo de luz que emerja del foco F y que se refleje en un elemento de parábola (cuya dirección coincide con la dirección de la tangente), partirá con una dirección paralela al eje Ox , es decir, paralela al eje de simetría de la parábola.

En esta propiedad de la parábola se basa la construcción de los telescopios reflectantes, como el inventado por Newton. Si construimos un espejo cóncavo cuya superficie sea un paraboloide de revolución —es decir, una superficie obtenida por rotación de una parábola alrededor de su eje de simetría—, todos los rayos

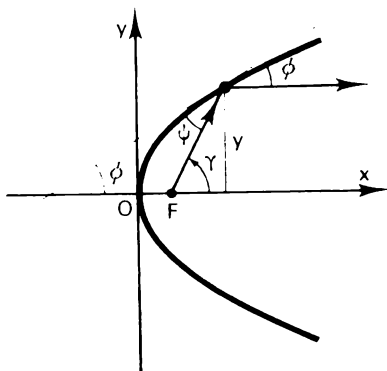


FIG. 25

de luz que partan de cualquier punto de un cuerpo celeste situado estrictamente en la dirección del «eje» del espejo se concentrarán, tras reflejarse en el espejo (Fig. 26), en un punto llamado foco. Los rayos que se originen en cualquier otro punto del cuerpo celeste y que no sean exactamente paralelos al eje del espejo se recogen en un punto próximo al foco. En el plano focal (el plano que pasa por el foco del espejo y es perpendicular a su eje) se obtiene la imagen invertida del astro; cuanto más lejos esté esta imagen del foco, más

difusa será, puesto que sólo los rayos exactamente paralelos al eje del espejo se recogen en un punto. La imagen obtenida puede ser observada mediante un microscopio especial (llamado ocular del telescopio), bien directamente o, bien —para no interrumpir los rayos de luz de la estrella con la cabeza— después de reflejarse en un pequeño espejo plano situado cerca del foco y a un ángulo de 45° .

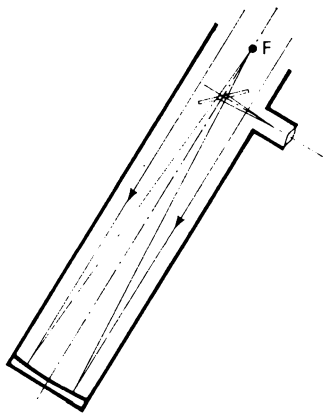


FIG. 26

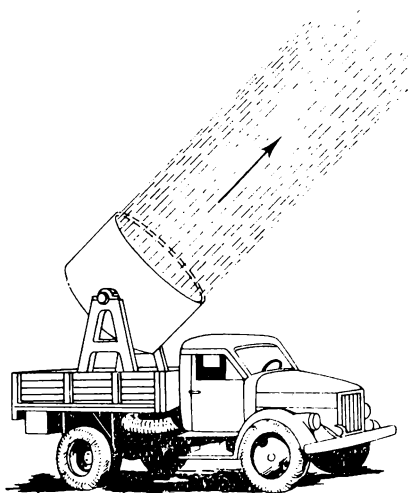


FIG. 27

Los grandes proyectores de seguimiento (Fig. 27) se basan en la misma propiedad de la parábola. En ellos, y al contrario de lo que ocurría en el caso anterior, se sitúa una fuente de luz muy intensa en el foco de un espejo parabólico, de modo que sus rayos salgan reflejados del espejo paralelamente a su eje. Los faros de los automóviles se construyen de forma parecida (Fig. 28).

En el caso de una elipse se demuestra fácilmente que los rayos que parten de uno de sus focos F_1 y se reflejan en la elipse se recogen en el otro foco F_2 (Fig. 29); y en la hipérbola los rayos que se originan en uno de sus focos F_1 se reflejan en ella como si partiesen del otro foco F_2 (Fig. 30).

Las directrices de la elipse y la hipérbola

Al igual que la parábola, la elipse y la hipérbola tienen directrices, pero dos en lugar de una. Si consideramos un foco y la directriz

«del mismo lado que él», entonces para todos los puntos M de la elipse tendremos $\varrho/d = \epsilon$, donde la constante ϵ es la excentricidad, que para una elipse es siempre menor que 1; y para todos los puntos de la correspondiente rama de la hipérbola, tenemos $\varrho/d = \epsilon$, donde ϵ es también la excentricidad, que para una hipérbola es siempre mayor que uno.



FIG. 28

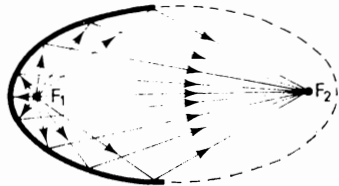


FIG. 29

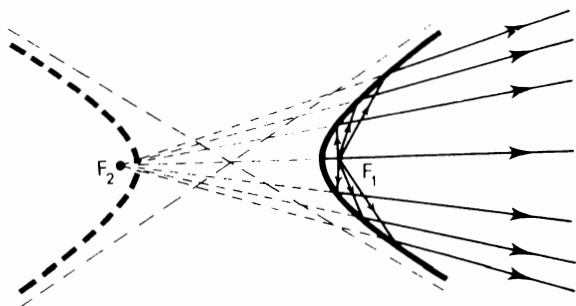


FIG. 30

Así, la elipse, la parábola y una rama de hipérbola son el lugar geométrico de los puntos del plano para los cuales el cociente de su distancia ϱ al foco y su distancia d a la directriz es constante (figuras 31 y 32).

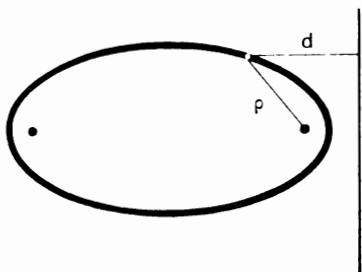


FIG. 31

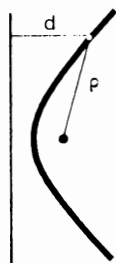


FIG. 32

Para la elipse esta constante es menor que 1; para la parábola es igual a 1, y para la hipérbola, mayor que 1. En este sentido la parábola es el caso «límite» o de «transición» entre la elipse y la hipérbola.

Secciones cónicas

Los griegos de la Antigüedad ya habían investigado con detalle las curvas obtenidas por intersección de un cono circular recto con un plano. Si el plano forma con el eje del cono un ángulo ϕ de 90° , es decir, si es perpendicular a él, la sección obtenida es una circunferencia. Es fácil demostrar que si el ángulo ϕ es menor de 90° , pero mayor que el ángulo α que forman las generatrices del cono con su eje, se obtiene una elipse. Si ϕ es igual a α , resulta una parábola, y si ϕ es menor que α , la sección obtenida es una hipérbola (Fig. 33).

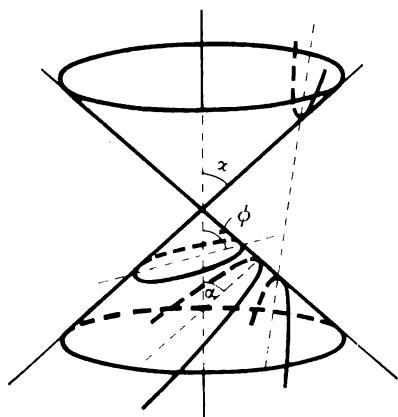


FIG. 33

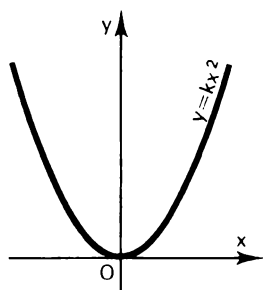


FIG. 34

La parábola como gráfica de la proporción cuadrática y la hipérbola como gráfica de la proporción inversa

Recordemos que la gráfica de la proporción cuadrática

$$y = kx^2$$

es una parábola (Fig. 34), y que la gráfica de la proporción inversa

$$y = \frac{k}{x} \quad \text{o} \quad xy = k$$

es una hipérbola (Fig. 35), como se demostrará más adelante. La hipérbola se definió anteriormente como la curva representada por la ecuación

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

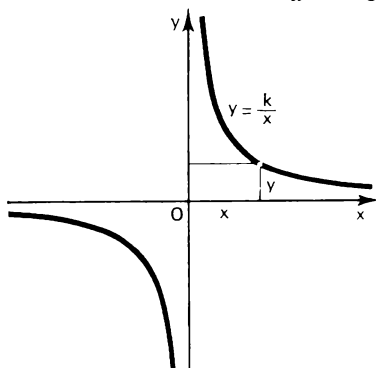


FIG. 35

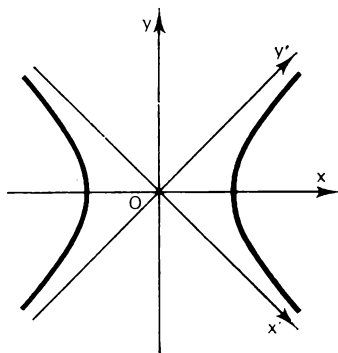


FIG. 36

En el caso especial $a = b$, la hipérbola (llamada entonces *equilátera*) viene a ser entre las hipérbolas lo que el círculo entre las elipses. En este caso, si se giran los ejes de coordenadas 45° (Fig. 36) la ecuación en las nuevas coordenadas (x', y') tendrá la forma

$$x'y' = k.$$

Hemos considerado hasta ahora tres curvas importantes de segundo orden: la elipse, la hipérbola y la parábola, y como definiciones hemos tomado las llamadas ecuaciones canónicas:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \text{ e } y^2 = 2px.$$

Pasemos ahora al estudio de la ecuación general de segundo grado en dos incógnitas, es decir, a discutir la clase de curvas que vienen representadas por esta ecuación.

§8. Reducción de la ecuación general de segundo grado a forma canónica

La primera exposición coherente de la geometría analítica, obra de Euler

Un hito importante en el desarrollo de la geometría analítica lo constituyó la aparición en 1748 del libro «Introducción al análisis», cuyo segundo volumen contenía —entre otras cosas relacionadas con

la teoría de funciones y otras ramas del análisis, y por primera vez en la historia de la matemática— la geometría analítica del plano, con un estudio detallado de las curvas de segundo orden; la exposición era muy parecida a la de los textos contemporáneos de geometría analítica, e iba acompañada también de un estudio de las curvas de orden superior. Esta fue la primera exposición de la geometría analítica en el sentido moderno de la palabra.

Reducción de una ecuación a forma canónica

La ecuación general de segundo grado³

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0$$

contiene seis términos, no tres o dos como en las ecuaciones canónicas de la elipse, hipérbola y parábola. Esto no significa que tal ecuación represente una curva más complicada, sino que el sistema de coordenadas elegido no es quizá el más conveniente para representarla. Si se elige un sistema conveniente de coordenadas cartesianas, una ecuación de segundo grado puede siempre reducirse a una de las siguientes formas canónicas:

- | | | | |
|----|---|---|--|
| 1. | $\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$ |  | Elipse |
| 2. | $\frac{x^2}{a^2} + \frac{y^2}{b^2} + 1 = 0$ |  | Elipse imaginaria |
| 3. | $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$ | | Punto (par de rectas imaginarias que se cortan en un punto real) |
| 4. | $\frac{x^2}{a^2} - \frac{y^2}{b^2} - 1 = 0$ |  | Hipérbola |
| 5. | $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$ |  | Par de rectas que se cortan |
| 6. | $y^2 - 2px = 0$ |  | Parábola |
| 7. | $x^2 - a^2 = 0$ |  | Par de rectas paralelas |
| 8. | $x^2 + a^2 = 0$ |  | Par de rectas paralelas imaginarias |
| 9. | $x^2 = 0$ |  | Par de rectas coincidentes |

donde a , b y p son distintas de cero.

³ Los coeficientes de xy , x , y se notarán, no por B , D , E , sino por $2B$, $2D$, $2E$, para simplificar las fórmulas subsiguientes.

Las ecuaciones 1, 4 y 6 de las formas canónicas que acabamos de enumerar son bien conocidas del lector; son las ecuaciones canónicas de la elipse, hipérbola y parábola. Dos de ellas, la 2 y la 8, no son satisfechas por ningún punto. En efecto, el cuadrado de un número real es siempre positivo o cero, por lo que en el primer miembro de la ecuación 2 la suma de los términos $x^2/a^2 + y^2/b^2$ no es nunca negativa, y puesto que también aparece el término $+1$, el primer miembro nunca puede ser igual a cero; análogamente, en la ecuación 8 el número x^2 es no negativo, y a^2 es positivo. De las mismas consideraciones se deduce que sólo el punto $(x = 0, y = 0)$, el origen, satisface la ecuación 3. La ecuación 5 puede escribirse como $(x/a - y/b)(x/a + y/b) = 0$, que la verifican aquellos puntos, y sólo aquellos, para los que una de las expresiones de primer grado $x/a - y/b$ o $x/a + y/b$ es igual a cero; así, la curva que representa es un par de rectas que se cortan. La ecuación 7 equivale análogamente a $(x - a)(x + a) = 0$; es decir, la curva correspondiente es un par de rectas paralelas: $x = a$ y $x = -a$. Por último, la curva 9 es un caso límite especial de la curva 7, cuando $a = 0$; es decir, es un par de rectas coincidentes.

Fórmulas de transformación de coordenadas

Para obtener el importante resultado de que antes hablamos sobre los tipos posibles de curvas de segundo orden, es necesario deducir antes las fórmulas que relacionan las coordenadas de los puntos en un cambio de sistema de coordenadas.

Sean x, y las coordenadas de un punto M respecto a los ejes Oxy . Traslademos estos ejes paralelamente a sí mismos hasta la posición $O'x'y'$, y sean ξ, η las coordenadas del nuevo origen O' respecto a los ejes antiguos. Es evidente (Fig. 37) que las nuevas coordenadas x', y'

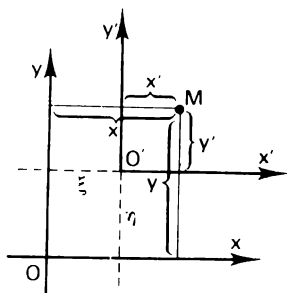


FIG. 37

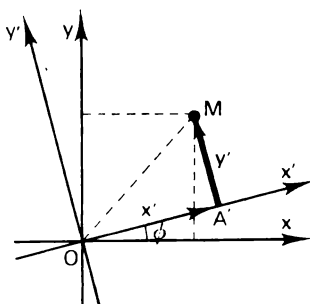


FIG. 38

del punto M están relacionadas con las antiguas x, y por medio de las fórmulas

$$x = x' + \xi,$$

$$y = y' + \eta,$$

que son las fórmulas de la llamada traslación paralela de ejes. Si giramos ahora los ejes Oxy en sentido contrario a las agujas del reloj, manteniendo fijo el origen de coordenadas, se deduce fácilmente (Fig. 38) que al proyectar la poligonal $OA'M$ formada por los segmentos de longitudes x', y' sobre los ejes Ox y Oy , respectivamente, obtenemos

$$x = x' \cos \phi - y' \sin \phi$$

$$y = x' \sin \phi + y' \cos \phi,$$

que son las fórmulas de transformación de coordenadas para una rotación del sistema de coordenadas rectangular.

Si nos dan la ecuación $F(x, y) = 0$ de una curva respecto a los ejes Oxy , y deseamos escribir la ecuación transformada de la misma curva respecto a los nuevos ejes $O'x'y'$, debemos reemplazar x e y en ecuación $F(x, y) = 0$ por sus expresiones en función de x' e y' , dadas por las fórmulas de transformación. Por ejemplo, en una traslación paralela de los ejes la ecuación transformada es

$$F(x' + \xi, y' + \eta) = 0,$$

y en una rotación

$$F(x' \cos \phi - y' \sin \phi, x' \sin \phi + y' \cos \phi) = 0.$$

Conviene observar que en una transformación de ejes el grado de la ecuación no varía. En efecto, el grado no puede aumentar, puesto que las fórmulas de transformación son de primer grado. Pero tampoco puede disminuir, puesto que entonces la transformación de coordenadas inversa (que también es de primer grado) tendría que aumentarlo.

Reducción de la ecuación general de segundo grado a una de las nueve formas canónicas

Vamos a demostrar ahora que cualquier ecuación de segundo grado en dos variables siempre se puede transformar (primero por rotación de ejes y luego por una traslación paralela) de modo tal que la ecuación referida a los nuevos ejes sea de una de las formas 1, 2, ..., 9.

En efecto, sea

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0 \quad (8)$$

la forma general de la ecuación de segundo grado.

Giremos ahora los ejes un ángulo ϕ , elegido del modo siguiente. Reemplazando x e y en la ecuación (8) por sus expresiones en función de las nuevas coordenadas (de acuerdo con las fórmulas para la rotación), se obtiene, después de reducir términos semejantes, que el coeficiente $2B'$ en la ecuación transformada

$$A'x'^2 + 2B'x'y' + C'y'^2 + 2D'x' + 2E'y' + F' = 0$$

es igual a

$$\begin{aligned} 2B' &= -2A \sin \phi \cos \phi + 2B(\cos^2 \phi - \sin^2 \phi) + 2C \sin \phi \cos \phi = \\ &= 2B \cos 2\phi - (A - C) \sin 2\phi. \end{aligned}$$

Iguálándolo a cero, se obtiene $2B \cos 2\phi = (A - C) \sin 2\phi$, por lo cual

$$\operatorname{ctg} 2\phi = \frac{A - C}{2B}.$$

Puesto que la cotangente varía de $-\infty$ a $+\infty$, siempre podemos encontrar un ángulo ϕ que verifique esta igualdad. Girando los ejes de este ángulo, se deduce que para los nuevos ejes $Ox' y'$ la ecuación de la curva, representada en los ejes primitivos por la ecuación (8), tiene la forma

$$A'x'^2 + C'y'^2 + 2D'x' + 2E'y' + F = 0, \quad (9)$$

es decir, no contiene el término en el producto de las coordenadas (F permanece invariable, puesto que las fórmulas de rotación no contienen términos constantes).

Una vez girados, traslademos ahora los ejes $Ox' y'$ paralelamente a sí mismos hasta la posición $O'' x'' y''$, y sean ξ', η' las coordenadas del nuevo origen O'' respecto a los ejes $Ox' y'$. La ecuación de la curva para estos nuevos ejes será:

$$A'(x'' + \xi')^2 + C'(y'' + \eta')^2 + 2D'(x'' + \xi') + 2E'(y'' + \eta') + F = 0. \quad (10)$$

Vamos a demostrar ahora que siempre se pueden seleccionar ξ' y η' (o lo que es lo mismo, que se pueden trasladar los ejes $Ox' y'$ paralelamente a sí mismos) de modo tal que la ecuación resultante para los ejes $O'' x'' y''$ tenga una de las formas canónicas 1, 2, . . . , 9.

Quitando todos los paréntesis en la ecuación (10) y reduciendo términos semejantes, se obtiene

$$A'x''^2 + C'y''^2 + 2(A'\xi' + D')x'' + 2(C'\eta' + E')y'' + F' = 0, \quad (10')$$

donde se ha notado por F' la suma de todos los términos constantes; su valor no interesa por el momento.

Consideremos tres casos posibles.

I. A' y C' no son simultáneamente nulos. En este caso, tomando $\xi' = -D'/A'$, $\eta' = -E'/C'$, se eliminan los términos en las primeras potencias de x'' e y'' y se obtiene una ecuación de la forma

$$A'x''^2 + C'y''^2 + F' = 0. \quad (I)$$

II. $A' \neq 0$, $C' = 0$, pero $E' \neq 0$. Poniendo $\xi' = -D'/A'$, $\eta' = 0$, es decir, $y'' = y'$, se obtiene la ecuación

$$A'x''^2 + 2E'y' + F' = 0,$$

o bien

$$A'x''^2 + 2E' \left(y' + \frac{F'}{2E'} \right) = 0.$$

Mediante una traslación paralela a lo largo del eje Oy' en una cantidad $\eta'' = -F'/2E'$, encontramos que $y' = y'' - F'/2E'$, es decir, $y' + F'/2E' = y''$, por lo que se obtiene la ecuación

$$A'x''^2 + 2E'y'' = 0. \quad (II)$$

Si es $A' = 0$, $C' \neq 0$, $D' \neq 0$, podemos simplemente intercambiar los papeles de x e y y obtener el mismo resultado.

III. $A' \neq 0$, $C' = 0$, $E' = 0$. Tomando otra vez $\xi' = -D'/A'$, $\eta' = 0$, obtenemos la ecuación

$$A'x''^2 + F' = 0. \quad (III)$$

En el caso $A' = 0$, $C' \neq 0$, $D' = 0$, podemos intercambiar de nuevo los papeles de x e y .

De este modo hemos agotado todas las posibilidades, pues A' y C' no pueden ser simultáneamente nulos, toda vez que entonces se habría reducido el grado de la ecuación, y hemos visto que en las transformaciones de coordenadas este grado no varía.

Así, con una elección apropiada del sistema de coordenadas, toda ecuación de segundo grado puede transformarse en una de las tres formas (I), (II), (III), llamadas «formas reducidas».

Supongamos que la ecuación es de la forma (I) (en este caso A' y C' son ambas distintas de cero). Si $F' \neq 0$, escribiendo la ecuación (I) en la forma

$$\frac{x''^2}{-F'/A'} + \frac{y''^2}{-F'/C'} - 1 = 0,$$

llegamos a una de las ecuaciones 1, 2 ó 4, según sea el signo de A' , C' , F' . Si el denominador de x''^2 es negativo y el de y''^2 positivo, hay también que intercambiar los ejes $O'' x''$ y $O'' y''$.

Si $F' = 0$, la ecuación (I) puede escribirse en la forma

$$\frac{x''^2}{1/A'} + \frac{y''^2}{1/C'} = 0,$$

y se llega a la ecuación 3 ó 5.

Si la ecuación tiene la forma (II) (en este caso ni A' ni E' son nulos), podemos escribirla como

$$x''^2 + \frac{2E'}{A'} y'' = 0,$$

y notando $-E'/A'$ por p e intercambiando los ejes $O'' x''$ y $O'' y''$, se obtiene la ecuación 6.

Finalmente, si la ecuación es de la forma (III) (con $A' \neq 0$), puede escribirse $x''^2 + F'/A' = 0$, y se obtiene una de las ecuaciones 7, 8 ó 9.

Este importante teorema sobre la posibilidad de reducir toda ecuación de segundo grado a una de las nueve formas canónicas fue desarrollado en detalle por Euler. Los razonamientos de Euler sólo difieren formalmente de los que acabamos de exponer.

§9. *Representación de fuerzas, velocidades y aceleraciones por ternas de números; teoría de vectores*

Un paso importante después de Euler lo dio Lagrange. En su «Mecánica analítica», publicada en 1788, Lagrange aritmetizó las fuerzas, las velocidades y las aceleraciones del mismo modo que Descartes aritmetizó los puntos. Esta idea, desarrollada por Lagrange en

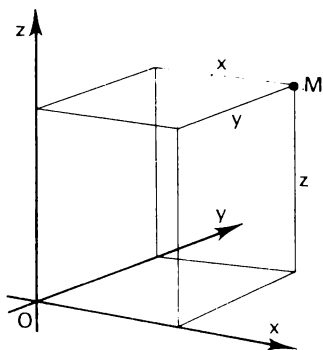


FIG. 39

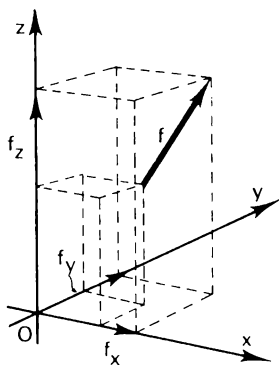


FIG. 40

su libro, tomó más tarde la forma de la llamada teoría de vectores y demostró ser de gran utilidad para la física, la mecánica y la tecnología.

Coordenadas rectangulares en el espacio

Señalemos, en primer lugar, que no fueron ni Descartes ni Newton quienes desarrollaron la geometría analítica en el espacio. Esto se hizo más tarde, en la primera mitad del siglo XVIII, gracias a la labor de Laguerre y Clairaut. Para especificar un punto M del espacio seleccionaron tres ejes mutuamente perpendiculares Ox , Oy y Oz y consideraron (Fig. 39) los valores numéricos de las distancias del punto M a los planos Oyz , Oxz y Oxy , tomadas con sus signos correspondientes, y llamadas *abscisa* x , *ordenada* y y *cota* z del punto M .

Aritmetización de las fuerzas, velocidades y aceleraciones, según Lagrange

Consideremos (Fig. 40) una fuerza f , que puede representarse en unidades convencionales por un segmento con punta de flecha, dotado de una longitud y una dirección específicas.

Lagrange señaló que esta fuerza f puede descomponerse en tres componentes f_x, f_y y f_z en la dirección de los ejes Ox, Oy y Oz , respectivamente; estas componentes, en tanto que segmentos dirigidos sobre los ejes, pueden definirse simplemente mediante un número positivo o negativo, según que la componente esté dirigida en la dirección del eje o en la opuesta. Así, podemos considerar, por ejemplo, la fuerza (2, 3, 4) o la fuerza (1, —2, 5), etc. En la composición de fuerzas según la ley del paralelogramo puede demostrarse fácilmente (se hará más adelante) que hay que sumar las componentes correspondientes. Por ejemplo, la suma de las fuerzas dadas es la fuerza

$$(2 + 1, 3 - 2, 4 + 5) = (3, 1, 9).$$

Lo mismo puede hacerse con las velocidades y las aceleraciones. En cualquier problema de mecánica, todas las ecuaciones que relacionan fuerzas, velocidades y aceleraciones pueden escribirse también como ecuaciones entre sus componentes, es decir, ecuaciones entre números. La ecuación mecánica se escribirá entonces en forma de tres ecuaciones: una para la x , otra para la y , y una tercera para la z .

Pero fue sólo cien años después de Lagrange cuando los matemáticos y físicos, bajo el influjo del desarrollo de la teoría de la electricidad, comenzaron a hacer uso abundante de la teoría general de los segmentos de longitud y dirección dadas. Tales segmentos se denominaron vectores.

La teoría de los vectores es de gran importancia en la mecánica, la física y la tecnología, y su aspecto algebraico, el álgebra vectorial (en contraste con el análisis vectorial) aparece en seguida como una parte esencial de la geometría analítica.

Algebra vectorial

Cualquier segmento dirigido (tanto si representa una fuerza, una velocidad, una aceleración u otra magnitud), es decir un segmento de longitud y dirección dadas, se llama *vector*. Dos vectores se dicen *iguales* si tienen la misma longitud y la misma dirección; es decir, en el concepto de «vector» sólo se tiene en cuenta su longitud y su dirección. Los vectores pueden ser sumados. Sean $a, b, \dots, d$ vectores dados. Eligiendo un punto como origen del vector a , tomamos su extremo como origen del vector b , etc. Obtenemos así un polígono vectorial $ab \dots d$ (Fig. 41). El vector m , cuyo punto inicial coincide con el punto inicial del primer vector a de este polígono, y cuyo

extremo coincide con el del último vector d , se llama *suma* de estos vectores

$$m = a + b + \cdots + d. \quad (11)$$

Es fácil demostrar que el vector m no depende del orden en el que se tomen los sumandos $a, b, \dots, d$.

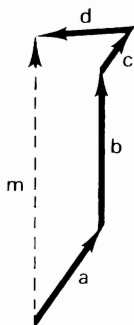


FIG. 41

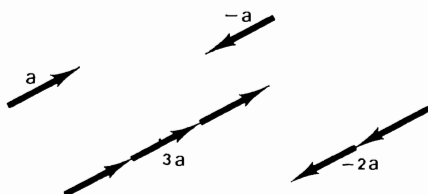


FIG. 42

El vector de igual longitud que a pero de dirección opuesta se llama vector *opuesto* al a y se nota por $-a$.

La sustracción del vector a se define como la suma de su vector opuesto.

En el cálculo vectorial los números reales ordinarios se denominan escalares. Dados un vector a (Fig. 42) y un escalar λ , llamaremos producto del vector a por el escalar (número) λ , es decir, λa , al vector cuya longitud es igual al producto de la longitud $|a|$ del vector a por el valor absoluto $|\lambda|$ del número λ , y cuya dirección es la misma que la de a si $\lambda > 0$, y la opuesta, si, $\lambda < 0$.

Consideremos un sistema de coordenadas cartesianas rectangulares $Oxyz$ y los vectores e_1, e_2, e_3 , de longitud unitaria y direcciones coincidentes con las direcciones positivas de los ejes Ox, Oy, Oz , respectivamente. Es evidente que cualquier punto dado M (Fig. 43) del espacio puede alcanzarse desde el origen O empalmando un cierto «número de veces» («número de veces» entero, fraccionario o irracional, positivo o negativo) el vector e_1 ; luego tantas «veces» el vector e_2 , y finalmente un cierto número de «veces» el vector e_3 . Está

claro que los números x, y, z , que muestran cuantas «veces» es necesario empalmar los vectores e_1, e_2 y e_3 , son sencillamente las coordenadas cartesianas del punto M .

Sea a un vector dado; si movemos un punto desde el origen de a hasta su extremo y descomponemos este movimiento en otros paralelos a los ejes Ox, Oy, Oz , y si para ello es necesario desplazar el

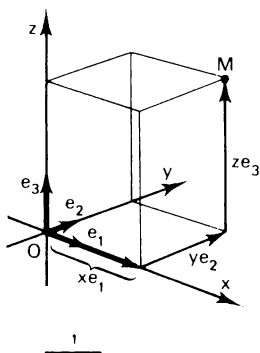


FIG. 43

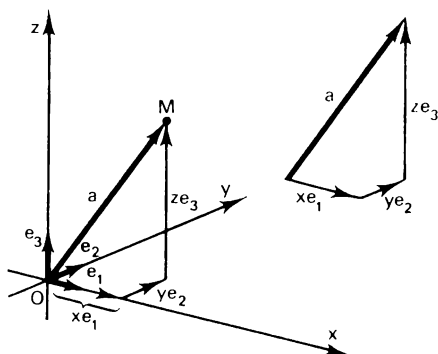


FIG. 44

punto una distancia xe_1 paralelamente al eje Ox , ye_2 paralelamente al eje Oy y ze_3 paralelamente al eje Oz , se tiene

$$a = xe_1 + ye_2 + ze_3. \quad (12)$$

Los números x, y, z se llaman *coordenadas* del vector a . Estas son evidentemente las coordenadas del punto extremo M de este vector, cuando su origen es el origen O del sistema de coordenadas (Fig. 44). De aquí se deduce inmediatamente que para sumar vectores deben sumarse sus coordenadas correspondientes, y para restarlos deben ser sustraídas. Si el primer vector «transporta» a lo largo del eje Ox una distancia xe_1 , y el segundo una distancia $x'e_1$, es evidente que su suma «transporta» a lo largo del eje Ox una distancia $(x + x')e_1$, etcétera (Fig. 45). De aquí se sigue también que para multiplicar un vector por un número, deben multiplicarse sus coordenadas por dicho número.

Producto escalar y sus propiedades

Dados dos vectores $\mathbf{a}$ y $\mathbf{b}$, el producto de sus longitudes por el coseno del ángulo que forman, $|\mathbf{a}| |\mathbf{b}| \cos \phi$, es un número llamado *producto escalar* de los vectores $\mathbf{a}$ y $\mathbf{b}$ y se nota por $\mathbf{ab}$ o $(\mathbf{ab})$. Sean x, y, z las coordenadas del vector $\mathbf{a}$, y $\bar{x}, \bar{y}, \bar{z}$ las coordenadas del vector $\mathbf{b}$; su producto escalar es entonces igual a

$$\mathbf{ab} = x\bar{x} + y\bar{y} + z\bar{z}, \quad (13)$$

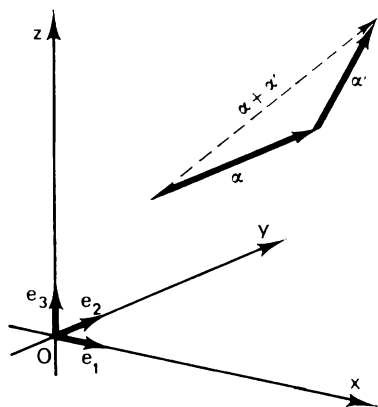


FIG. 45

es decir, a la suma de los productos de sus coordenadas correspondientes.

Este importante resultado puede demostrarse como sigue. Hagamos primero las siguientes observaciones:

1) Si se multiplica uno de los vectores de un producto escalar, por ejemplo $\mathbf{a}$, por un número λ , el resultado es evidentemente el mismo que si se multiplica el producto escalar por el mismo número, es decir,

$$(\mathbf{a}\lambda)\mathbf{b} = \lambda(\mathbf{ab}).$$

2) El producto escalar es distributivo, es decir, si $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$, entonces $\mathbf{ab} = \mathbf{a}_1\mathbf{b} + \mathbf{a}_2\mathbf{b}$.

En efecto, el primer miembro de esta igualdad es igual al producto de la longitud del vector $\mathbf{b}$ por el valor numérico de la proyección del vector $\mathbf{a}$ sobre el eje del vector $\mathbf{b}$ (Fig. 46), y el segundo miembro es igual al producto de la longitud de $\mathbf{b}$ por la suma de los valores numéricos de las proyecciones de los vectores $\mathbf{a}_1$ y $\mathbf{a}_2$ sobre el eje de $\mathbf{b}$. Pero $\text{proyec } \mathbf{a} = \text{proyec } \mathbf{a}_1 + \text{proyec } \mathbf{a}_2$, lo que demuestra la igualdad.

Consideremos ahora dos vectores $\mathbf{a}$ y $\mathbf{b}$ cuyas descomposiciones según los vectores $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ son $\mathbf{a} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$, $\mathbf{b} = \bar{x}\mathbf{e}_1 + \bar{y}\mathbf{e}_2 + \bar{z}\mathbf{e}_3$, por lo que

$$\mathbf{ab} = (x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3)(\bar{x}\mathbf{e}_1 + \bar{y}\mathbf{e}_2 + \bar{z}\mathbf{e}_3).$$

En virtud de la distribución del producto escalar, las dos sumas de vectores entre paréntesis pueden ser multiplicadas como polinomios, y, en virtud de (1), los factores escalares de cada uno de los términos pueden sacarse fuera de los paréntesis, por lo que

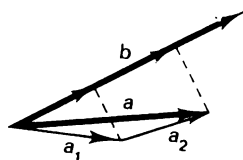
$$\begin{aligned} \mathbf{ab} = & x\bar{x}\mathbf{e}_1\mathbf{e}_1 + x\bar{y}\mathbf{e}_1\mathbf{e}_2 + x\bar{z}\mathbf{e}_1\mathbf{e}_3 + y\bar{x}\mathbf{e}_2\mathbf{e}_1 + y\bar{y}\mathbf{e}_2\mathbf{e}_2 + \\ & + y\bar{z}\mathbf{e}_2\mathbf{e}_3 + z\bar{x}\mathbf{e}_3\mathbf{e}_1 + z\bar{y}\mathbf{e}_3\mathbf{e}_2 + z\bar{z}\mathbf{e}_3\mathbf{e}_3. \end{aligned}$$

Pero

$$|\mathbf{e}_1| = |\mathbf{e}_2| = |\mathbf{e}_3| = 1, \quad \cos 0^\circ = 1 \quad \text{y} \quad \cos 90^\circ = 0.$$

Por tanto,

$$\begin{aligned} \mathbf{e}_1\mathbf{e}_1 &= 1, & \mathbf{e}_1\mathbf{e}_2 &= 0, & \mathbf{e}_1\mathbf{e}_3 &= 0, \\ \mathbf{e}_2\mathbf{e}_1 &= 0, & \mathbf{e}_2\mathbf{e}_2 &= 1, & \mathbf{e}_2\mathbf{e}_3 &= 0, \\ \mathbf{e}_3\mathbf{e}_1 &= 0, & \mathbf{e}_3\mathbf{e}_2 &= 0, & \mathbf{e}_3\mathbf{e}_3 &= 1. \end{aligned}$$



Así,

$$\mathbf{ab} = x\bar{x} + y\bar{y} + z\bar{z}. \quad (14)$$

FIG. 46

Observemos, en particular, que, si los vectores $\mathbf{a}$ y $\mathbf{b}$ son mutuamente perpendiculares, entonces $\phi = 90^\circ$ y $\cos \phi = 0$. Por tanto la igualdad

$$x\bar{x} + y\bar{y} + z\bar{z} = 0 \quad (15)$$

sirve como condición (fácilmente verificable) de perpendicularidad de los vectores $\mathbf{a}$ y $\mathbf{b}$.

Angulo formado por dos direcciones

Consideremos una dirección caracterizada por los ángulos α, β, γ que forma con los ejes de coordenadas. Dibujemos la recta que, con esta dirección, pasa por el origen del sistema de coordenadas, y tomemos sobre ella, a partir del origen, un segmento OA de longitud unidad (Fig. 47). En este caso las coordenadas del punto A , es decir, las coordenadas del vector $\vec{OA}$ son exactamente $\cos \alpha, \cos \beta$ y $\cos \gamma$. Si tenemos una segunda dirección dada por los ángulos $\bar{\alpha}, \bar{\beta}, \bar{\gamma}$, entonces el vector análogo $\vec{OB}$ para esta segunda dirección tiene por coordenadas $\cos \bar{\alpha}, \cos \bar{\beta}, \cos \bar{\gamma}$ (Fig. 48). Sea ϕ el ángulo formado

por estos vectores; su producto escalar es entonces igual a $1 \cdot 1 \cos \phi$, y tenemos que:

$$\cos \phi = \cos \alpha \cos \bar{\alpha} + \cos \beta \cos \bar{\beta} + \cos \gamma \cos \bar{\gamma}, \quad (16)$$

fórmula muy importante para obtener el coseno del ángulo formado por dos direcciones.

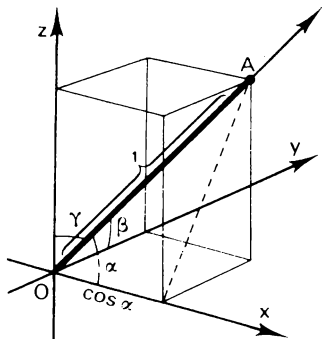


FIG. 47

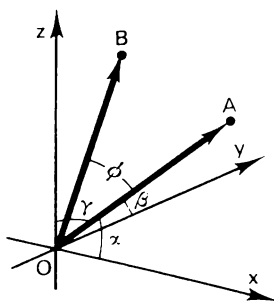


FIG. 48

§10. Geometría analítica del espacio; ecuaciones de una superficie y de una curva en el espacio

Si en la ecuación $z = f(x, y)$ consideramos x e y como la abscisa y la ordenada de un punto y z como su cota, esta ecuación representa una superficie P , que puede obtenerse elevando perpendiculares de longitud z en los puntos (x, y) del plano Oxy . El lugar geométrico de los extremos de estas perpendiculares es la superficie P representada por esta ecuación. Si la ecuación que relaciona x, y, z no está ya resuelta con respecto a z , se puede resolver y construir después la superficie P . En general, en geometría analítica el conjunto de todos los puntos del espacio cuyas coordenadas x, y, z satisfacen una ecuación dada (Fig. 49) en tres variables, x, y, z , se dice que es la superficie representada por la ecuación.

Como ya indicamos en el capítulo 2, una función de dos variables $f(x, y)$ no sólo puede representar una superficie P , sino también su sistema de curvas de nivel, es decir, curvas del plano Oxy en cada una de las cuales la función $f(x, y)$ tiene un valor constante. Este sistema de curvas no es otra cosa que el mapa topográfico de la superficie P sobre el plano Oxy .

Ejemplo

La ecuación $xy = z$ tiene, por ejemplo, las curvas de nivel: $\dots$, $xy = -3$, $xy = -2$, $xy = -1$, $xy = 0$, $xy = 1$, $xy = 2$, $xy = 3$, $\dots$. Todas ellas son hipérbolas (Fig. 50) excepto $xy = 0$,

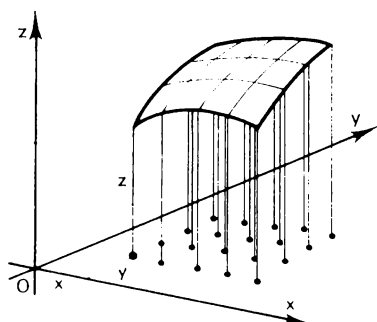


FIG. 49

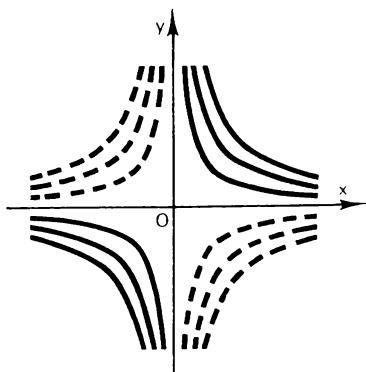


FIG. 50

que representa los dos ejes de coordenadas. Lo que se obtiene es una superficie en forma de silla de montar (Fig. 51) (llamada paraboloides hiperbólico).

Para definir una curva en el espacio, podemos dar las ecuaciones de dos superficies cualesquiera P y Q que se cortan según la curva. Por ejemplo, el sistema

$$\begin{aligned} xy &= z, \\ x^2 + y^2 &= 1 \end{aligned}$$

define una curva en el espacio (Fig. 52). La ecuación $xy = z$ determina el paraboloides hiperbólico ya citado; y la ecuación $x^2 + y^2 = 1$ determina un cilindro circular de radio unidad, cuyo eje es el eje Oz . El sistema de ecuaciones define, por tanto, la curva intersección del paraboloides con el cilindro, curva que se ha representado en la figura 52.

Si en este sistema se elige arbitrariamente una de las incógnitas, por ejemplo la x , y se resuelve respecto a y y z , se irán obteniendo las coordenadas x, y, z , de los distintos puntos de la curva.

Ecuaciones de un plano y de una recta

Puede demostrarse que toda ecuación de primer grado con tres variables

$$Ax + By + Cz + D = 0$$

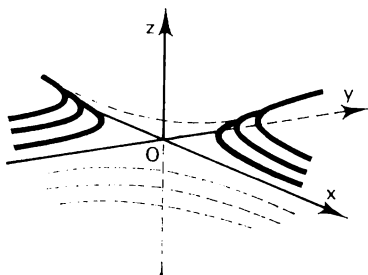


FIG. 51

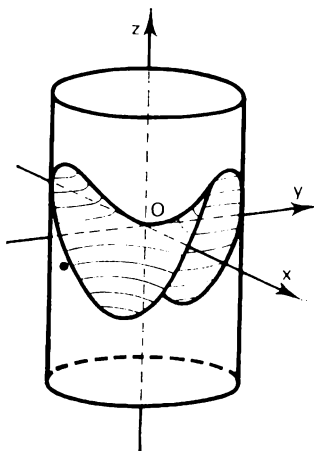


FIG. 52

representa un plano, y recíprocamente. De lo dicho anteriormente se desprende que una recta puede venir definida mediante un sistema de dos ecuaciones de este tipo:

$$A_1x + B_1y + C_1z + D_1 = 0,$$

$$A_2x + B_2y + C_2z + D_2 = 0,$$

es decir, como la línea intersección de dos planos.

La ecuación general de segundo grado en tres variables y sus 17 formas canónicas

La ecuación general de segundo grado en tres variables

$$A_1x^2 + A_2y^2 + A_3z^2 + 2B_1yz + 2B_2xz + 2B_3xy + \\ + 2C_1x + 2C_2y + 2C_3z + D = 0, \quad (17)$$

contiene diez términos. Análogamente a lo que se hizo anteriormente para una ecuación con dos variables, puede demostrarse que por una

rotación conveniente del sistema de coordenadas alrededor del origen la ecuación (17) puede reducirse a la forma

$$A_1'x'^2 + A_2'y'^2 + A_3'z'^2 + 2C_1'x' + 2C_2'y' + 2C_3'z' + D = 0, \quad (18)$$

es decir, se pueden eliminar los términos que contienen los productos cruzados. Sin embargo, la demostración de la posibilidad de tal simplificación es considerablemente más difícil que en el caso del plano. La dificultad estriba en que una rotación en el plano alrededor de un punto viene definida por un único ángulo ϕ , que elegimos convenientemente, mientras que en el espacio la rotación de un cuerpo alrededor de un punto fijo viene definida por tres ángulos independientes (ángulos de Euler) ϕ, θ, ψ de un modo complicado. Esto hace que los puntos cruzados tengan que eliminarse de un modo indirecto (ver capítulo 15 sobre la teoría de reducción de una forma cuadrática a suma de cuadrados por transformaciones ortogonales). Luego, como en el caso del plano, se hace una traslación paralela de los ejes y se simplifica la ecuación; la ecuación (18) tomará entonces una de las formas canónicas siguientes:

- | | |
|--|--|
| 1. $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = 0$ |  Elipsoide |
| 2. $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} + 1 = 0$ |  Elipsoide imaginario |
| 3. $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} - 1 = 0$ |  Hiperboloide de una hoja |
| 4. $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} + 1 = 0$ |  Hiperboloide de dos hojas |
| 5. $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$ |  Cono de segundo orden |
| 6. $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 0$ |  Cono de segundo orden imaginario |
| 7. $\frac{x^2}{a^2} + \frac{y^2}{b^2} - 2cz = 0$ |  Paraboloide elíptico |

- | | | |
|--|---|---|
| 8. $\frac{x^2}{a^2} - \frac{y^2}{b^2} - 2cz = 0$ |  | Paraboloide hiperbólico |
| 9. $\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$ |  | Cilindro elíptico |
| 10. $\frac{x^2}{a^2} + \frac{y^2}{b^2} + 1 = 0$ |  | Cilindro elíptico imaginario |
| 11. $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$ |  | Par de planos imaginarios que se cortan |
| 12. $\frac{x^2}{a^2} - \frac{y^2}{b^2} - 1 = 0$ |  | Cilindro hiperbólico |
| 13. $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$ |  | Par de planos que se cortan |
| 14. $y^2 - 2px = 0$ |  | Cilindro parabólico |
| 15. $x^2 - a^2 = 0$ |  | Par de planos paralelos |
| 16. $x^2 + a^2 = 0$ |  | Par de planos paralelos imaginarios |
| 17. $x^2 = 0$ |  | Par de planos coincidentes |

Las ecuaciones canónicas 9-17 no contienen términos en z y representan exactamente las ecuaciones canónicas de curvas de segundo orden en el plano Oxy . En el espacio estas ecuaciones representan cilindros cuyas directrices son las correspondientes curvas de segundo orden en el plano Oxy cuyas generatrices son paralelas al eje Oz . En efecto, si una de estas ecuaciones se verifica para un punto de coordenadas $(x_1, y_1, 0)$, también se verificará para un punto de coordenadas (x_1, y_1, z) cualquiera que sea z , puesto que no hay términos en z en la ecuación.

Entre las ecuaciones 1-8, como puede verse fácilmente, la ecuación 2 no se satisface para ningún punto con x, y, z reales; y la

ecuación 6 se satisface sólo para el punto $(0, 0, 0)$, es decir, el origen. Así, pues, quedan únicamente por estudiar las seis ecuaciones 1, 3, 4, 5, 7 y 8.

Elipsoide

Comparemos las superficies representadas por las ecuaciones $x^2/a^2 + y^2/b^2 + z^2/c^2 - 1 = 0$ y $x^2 + y^2 + z^2 - 1 = 0$. La segunda es evidentemente la ecuación de una esfera C de centro el origen de coordenadas y radio unidad, puesto que $x^2 + y^2 + z^2$ es el cuadrado de la distancia del punto (x, y, z) al origen O . Si (x, y, z) es un punto de la esfera, es decir, si satisface la segunda, ecuación, (ax, by, cz) será un punto cuyas coordenadas satisfacen la primera ecuación. La superficie representada por la primera ecuación se obtiene así a partir de la esfera C , reemplazando todas las coordenadas x de los puntos de la esfera por ax , las coordenadas y por by , y las z por cz , es decir, deformando uniformemente la esfera C a partir de los planos Oyz , Oxz , Oxy con coeficientes de dilatación a , b y c , respectivamente. Esta superficie se llama elipsoide (Fig. 53).

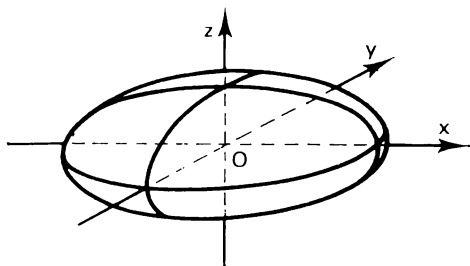


FIG. 53

Hiperboloides y el cono de segundo orden

Consideremos las ecuaciones 3, 4 y 5, es decir, la ecuación de la forma

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = \delta, \quad (19)$$

donde $\delta = 1, -1$ ó 0 . Comparémosla con la ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} - \frac{z^2}{c^2} = \delta, \quad (20)$$

en la cual el denominador de y^2 es también a^2 y no b^2 , como en la ecuación (19). Como antes, observamos que la superficie (19) se obtiene

a partir de la superficie (20) mediante una dilatación desde el plano Oxz con coeficiente b/a .

Veamos ahora qué superficie representa la ecuación (20). Tomemos un plano $z = h$ perpendicular al eje Oz y examinemos su inter-

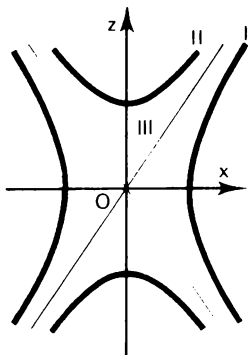


FIG. 54

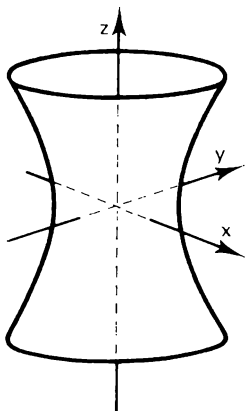


FIG. 55

sección con la superficie (20). Sustituyendo $z = h$ en la ecuación (20), obtenemos

$$x^2 + y^2 = a^2 \left(\delta + \frac{h^2}{c^2} \right).$$

Si $\delta + h^2/c^2$ es positivo, esta ecuación, junto con $z = h$, representa una circunferencia situada en el plano $z = h$, y con centro en el eje Oz . Si $\delta + h^2/c^2$ es negativo (lo que sólo puede ocurrir cuando $\delta = -1$ y h^2 sea pequeño), el plano $z = h$ no corta a la superficie (20), puesto que la suma de los cuadrados $x^2 + y^2$ no puede ser un número negativo.

La superficie (20) consta así de circunferencias situadas en planos perpendiculares al eje Oz y con centros sobre dicho eje. Así, pues, la superficie (20) es una superficie de revolución alrededor del eje Oz . Si la intersectamos con un plano que pase por el eje Oz , obtenemos su «meridiano», es decir, una curva, situada en un plano que pasa por el eje, cuya revolución alrededor de éste genera la superficie.

Si cortamos la superficie (20) con el plano coordenado Oxz , es decir, el plano $y = 0$ (Fig. 54), para lo cual sustituimos $y = 0$ en la ecuación (20), obtenemos la ecuación del meridiano $x^2/a^2 - z^2/c^2 = \delta$. Si $\delta = 1$, se trata de la hipérbola tipo *I*; si $\delta = -1$, de la hipérbola *II*; y si $\delta = 0$, del par de rectas que se cortan *III*. Por revolución alrededor del eje z engendran, respectivamente, el llamado hiperboloide de revolución de una hoja (Fig. 55), el hiperboloide de dos hojas (Fig. 56) y un cono circular recto (Fig. 57).

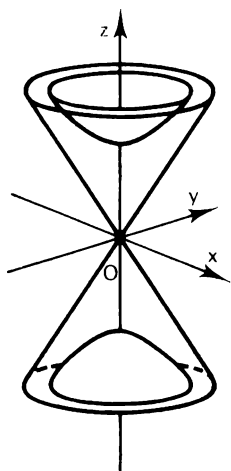


FIG. 56

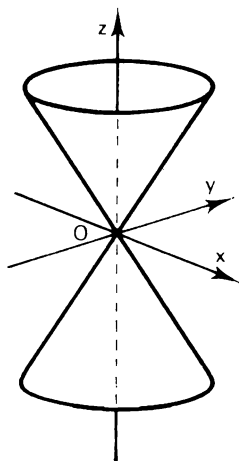


FIG. 57

El hiperboloide general de una hoja, el de dos hojas y el cono de segundo orden (3, 4 y 5) se obtienen a partir de estas superficies de revolución por una dilatación a partir del plano Oxz con coeficiente igual a b/a .

Paraboloides

Quedan sólo por estudiar las ecuaciones 7 y 8. Comparemos la primera de estas $x^2/a^2 + y^2/b^2 = 2cz$ con la ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} = 2cz,$$

que estudiaremos del mismo modo que la anterior. Esta ecuación representa la superficie obtenida por revolución de la parábola

$x^2 = 2a^2cz$ alrededor del eje Oz : lo que llamamos un paraboloide de revolución (Fig. 58), del que ya hemos hablado anteriormente al referirnos a los espejos parabólicos. El paraboloide elíptico general 7 se obtiene a partir del paraboloide de revolución mediante una dilatación desde el plano Oxz .

La superficie 8 hay que estudiarla de un modo diferente, examinando sus intersecciones con los planos $z = h$; las intersecciones son

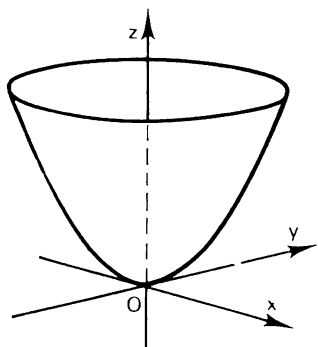


FIG. 58

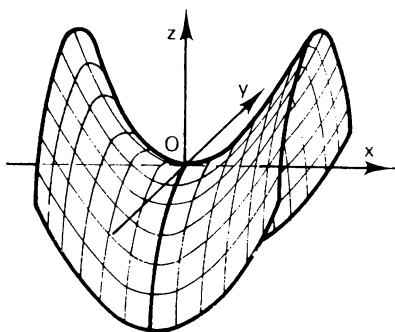


FIG. 59

hipérbolas. El mapa de contorno de la superficie 8 está representado en la figura 50; aunque con una posición distinta de los ejes coordenados, esta superficie ya se consideró en la figura 51. Tiene forma de silla de montar, como se ve en la figura 59, y se llama paraboloide hiperbólico. Sus intersecciones con los planos paralelos al plano Oxz resultan ser parábolas idénticas. El mismo resultado se obtiene al cortarlo por planos paralelos al Oyz .

Generatrices rectilíneas del hiperboloide de una hoja

Es muy curioso el hecho, no del todo obvio, de que el hiperboloide de una hoja y el paraboloide hiperbólico pueden obtenerse, al igual que el cono y el cilindro, por el movimiento de una línea recta. En el caso del hiperboloide basta demostrarlo para un hiperboloide de revolución de una hoja $x^2/a^2 + y^2/b^2 - z^2/c^2 = 1$, puesto que el hiperboloide general de una hoja se obtiene por dilatación uniforme a partir del plano Oxz , y en una dilatación toda recta se transforma

en una recta. Cortemos el hiperboloide de revolución con el plano $y = a$ paralelo al plano Oxz . Sustituyendo $y = a$ obtenemos

$$\frac{x^2}{a^2} + \frac{a^2}{a^2} - \frac{z^2}{c^2} = 1, \text{ o bien } \frac{x^2}{a^2} - \frac{z^2}{c^2} = 0.$$

Pero esta ecuación, junto con $y = a$, da en el plano $y = a$ un par de rectas que se cortan $x/a - z/c = 0$ y $x/a + z/c = 0$.

Se comprueba así que el hiperboloide contiene un par de rectas que se cortan. Si se gira el hiperboloide alrededor del eje Oz , entonces cada una de estas dos rectas genera todo el hiperboloide (Fig. 60).

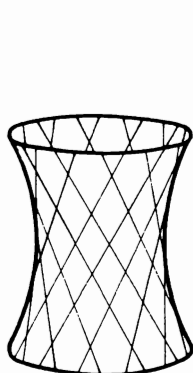


FIG. 60

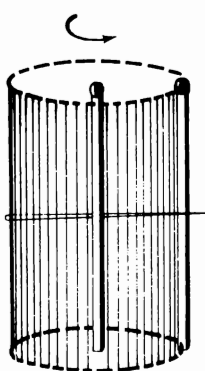


FIG. 61

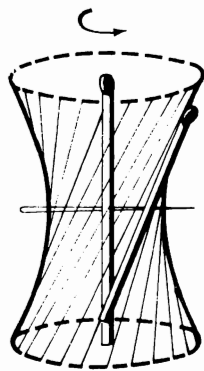


FIG. 62

Es fácil demostrar que: 1) dos rectas cualesquiera de una y la misma familia de rectas así obtenidas no se encuentran en el mismo plano (son rectas que se cruzan); 2) cualquier recta de una de estas familias corta a todas las rectas de la otra (excepto a su opuesta, que es paralela a ella); 3) tres rectas cualesquiera de la misma familia nunca son paralelas a un mismo plano.

Con dos cerillas y una aguja es fácil obtener una representación del hiperboloide de revolución de una hoja. Se atraviesa una de las cerillas con la aguja, y en la punta se pincha la otra cerilla paralelamente a la primera. Si giramos ahora todo el conjunto alrededor de la primera cerilla como eje, la segunda recorrerá la superficie de un cilindro (Fig. 61); pero, si la segunda cerilla no es paralela a la pri-

mera, en una revolución completa describirá la superficie de un hiperboloide de revolución de una hoja, como puede visualizarse fácilmente si la rotación es rápida (Fig. 62).

Resumen del estudio de la ecuación de segundo grado

Aunque la ecuación general de segundo grado con tres variables puede representar 17 superficies distintas, no es difícil recordarlas. Las 9 últimas son cilindros sobre las 9 curvas posibles de segundo orden, mientras que las 8 primeras se dividen en cuatro pares: dos elipsoides (real e imaginario), dos hiperboloides (de una hoja y de dos hojas), dos conos de segundo orden (real e imaginario) y dos paraboloides (elíptico e hiperbólico). Todas estas superficies juegan un papel esencial en la mecánica, la física y la tecnología (elipsoide de inercia, elipsoide de elasticidad, hiperboloide en la transformación de Lorentz en física, paraboloide de revolución para los espejos parabólicos, etc.).

§11. Transformaciones afín y ortogonal

La siguiente etapa importante en el desarrollo de la geometría analítica fue la introducción (tanto en ella como en toda la geometría) de la teoría de las transformaciones. Vamos a estudiarla en detalle.

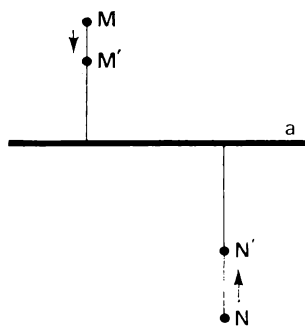


FIG. 63

«Contracción» del plano a una línea

Consideremos una de las transformaciones más sencillas del plano, la llamada «contracción» uniforme de coeficiente k hacia una recta. Sea a una recta dada del plano y k un coeficiente positivo, por ejemplo, $k = 2/3$. Todos los puntos de la recta a quedan fijos, y todo punto M no perteneciente a esta recta se transforma en el punto M' , tal que M' se encuentra en el mismo semiplano que M , sobre la perpendicular de M a a y a una distancia de a igual a $2/3$ de la distancia de M a a . Si el

coeficiente k , como en este caso, es menor que la unidad, tenemos una contracción propia del plano a la recta; pero si k es mayor que la unidad, tenemos una dilatación del plano a partir de la recta; por

conveniencia hablaremos en este y otros casos de «contracción», colocando la palabra «contracción» entre comillas.

El punto o figura a transformar se llama preimagen, y la figura transformada su imagen. El punto M' , por ejemplo, es la imagen del punto M (Fig. 63).

Vamos a demostrar que en una «contracción» uniforme del plano a una recta, cualquier recta del plano se transforma en otra recta. Supongamos que el plano se «contrae» a una recta a de él, con coeficiente de «contracción» k . Sea b cualquier recta del plano, O su punto de intersección con la recta a , B otro punto arbitrario de b , y BA la perpendicular a la recta a desde el punto B (Fig. 64). En la

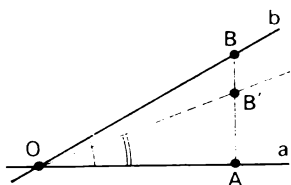


FIG. 64

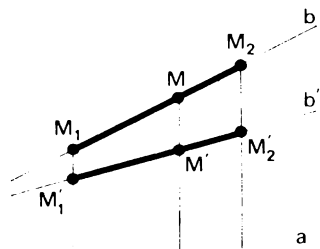


FIG. 65

«contracción» el punto B se transforma en el B' , situado sobre esta perpendicular de modo que $B'A = k \cdot BA$. Por tanto, la tangente del ángulo $B'OA$ será igual a $AB'/OA = k \cdot AB/OA$, es decir, será igual a k veces la tangente del ángulo que la recta b forma con la recta a , e idéntica para todas las imágenes de los puntos de b . Todos los puntos B' se encontrarán por tanto en una misma recta, que pasa por el punto O y forma con la recta a un ángulo con dicha tangente.

En una «contracción» las rectas paralelas siguen siendo paralelas. En efecto, si las tangentes de los ángulos que las rectas b y c forman con la a son iguales, las tangentes de los ángulos que las imágenes b' y c' forman con a difieren sólo en un factor k , es decir, son iguales entre sí, lo que significa que las rectas b' y c' siguen siendo paralelas.

En una «contracción» a una recta, cualquier segmento rectilíneo se contrae (o expande) uniformemente (aunque en distinto grado para segmentos de distintas direcciones). Cuando hablamos de contracción «uniforme», queremos decir que el punto medio del segmento sigue siendo el punto medio, el punto situado a un tercio del extremo

sigue estando situado a un tercio del extremo, etc., es decir, el segmento se contrae uniformemente en toda su longitud. En efecto, cualquiera que sea la razón en que el punto M divide al segmento M_1M_2 , su imagen M' dividirá a $M'_1M'_2$ en la misma razón, puesto que las rectas paralelas (en este caso perpendiculares a la recta a) cortan a las rectas (en este caso b y b') en partes proporcionales (Fig. 65).

La elipse como resultado de la «contracción» de la circunferencia

Consideremos una circunferencia de centro el origen y radio a . Por el teorema de Pitágoras su ecuación es $x^2 + \bar{y}^2 = a^2$, donde hemos escrito $\bar{y}$ en lugar de y porque la y se necesitará más tarde. Veamos en qué se convierte esta circunferencia si «contraemos» el plano al eje Ox con un coeficiente igual a b/a (Fig. 66). Después de esta «contracción» los valores x de todos los puntos permanecen idénticos, mientras que los $\bar{y}$ se transformarán en $y = \bar{y}(b/a)$, es decir, $\bar{y} = (a/b)y$. Sustituyendo $\bar{y}$ en la ecuación de la circunferencia, tendremos:

$$x^2 + \frac{a^2}{b^2} y^2 = a^2$$

$$\text{o bien } \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

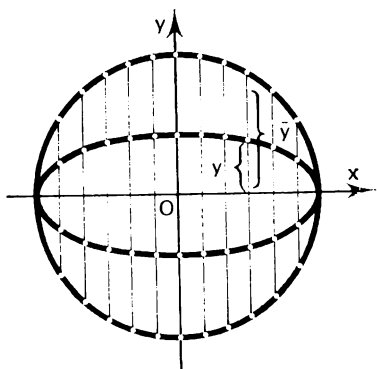


FIG. 66

que es la ecuación, en el mismo sistema de coordenadas, de la curva obtenida a partir de la circunferencia dada por contracción al eje Ox . Como vemos, se obtiene una elipse. Hemos demostrado, pues, que una elipse es el resultado de la «contracción» de una circunferencia.

Del hecho de que una elipse es una «contracción» de una circunferencia se deducen directamente muchas propiedades de la elipse; por ejemplo, la propiedad de los diámetros (mencionada anteriormente), que dice que los puntos medios de un conjunto de secantes paralelas de una elipse se encuentran alineados (ver Fig. 12), puede demostrarse del modo siguiente. Realizamos la dilatación inversa, pasando de la elipse a la circunferencia. En esta dilatación las secan-

tes paralelas de la elipse se transforman en cuerdas paralelas de la circunferencia, y los puntos medios de aquéllas en puntos medios de estas cuerdas. Pero los puntos medios de las cuerdas paralelas de una circunferencia se encuentran sobre un diámetro, es decir, sobre una línea recta, y así los puntos medios de las secantes paralelas de la elipse también estarán alineados. Es decir, se encontrarán sobre la recta en que se transforma el diámetro de la circunferencia como resultado de la «contracción» que transforma la circunferencia en elipse.

Existe otra aplicación de la teoría de la «contracción». Puesto que cualquier banda vertical del círculo no modifica su anchura al contraerse al eje Ox y su longitud se multiplica por b/a , el área de esta banda después de la contracción es igual a la inicial multiplicada por b/a ; pero, como el área del círculo es igual a πa^2 , el área de la elipse correspondiente será igual a $\pi a^2(b/a) = \pi ab$.

Ejemplo de solución de un problema más complicado

Dada una elipse se pide encontrar el triángulo de menor área circunscrito a esta elipse. Resolvamos primero el problema para una circunferencia. Demostraremos que, en el caso de una circunferencia, el triángulo es equilátero. En efecto, supongamos que el triángulo circunscrito no es equilátero; es decir, el más pequeño de sus ángulos (notado por B) es menor que 60° , y el mayor $C > 60^\circ$. Si entonces, sin modificar el ángulo A , movemos el lado BC hacia posición $B_0 C_0$ (Fig. 67), desplazando el vértice B hacia el A hasta que uno de los ángulos B_0 o C_0 sea igual a 60° , obtenemos el triángulo circunscrito $AB_0 C_0$, cuya área es más pequeña que la de ABC , puesto que ⁴ $OC < OB$, $OC_0 \leq OB_0$ y por tanto el área desechada OBB_0 es mayor que la añadida OCC_0 . Si el triángulo obtenido no es equilátero, repitiendo el proceso anterior reducimos su área todavía más hasta llegar a un triángulo equilátero. Por consiguiente, el área de todo triángulo no equilátero circunscrito a una circunferencia dada es mayor que el área del triángulo equilátero circunscrito a esa misma circunferencia.

Volvamos a la elipse. Dilatémosla desde su eje mayor, convirtiéndola de nuevo en la circunferencia a partir de la cual se obtuvo por «contracción». En esta dilatación (Fig. 68): 1) todos los triángulos circunscritos a la elipse se transforman en triángulos circunscritos a la circunferencia resultante; 2) las áreas de todas las figuras,

⁴ Como puede demostrarse fácilmente.

y en particular las de estos triángulos, se incrementan en la misma razón. Por tanto vemos que, de los triángulos circunscritos a la elipse dada, los de menor área serán aquellos que se convierten en triángulos equiláteros circunscritos a la circunferencia. Hay un número infinito de tales triángulos; cada uno de ellos tiene su centro de gravedad en el centro de la elipse y los puntos de tangencia ocupan el punto medio de sus lados. Cualquiera de estos triángulos puede

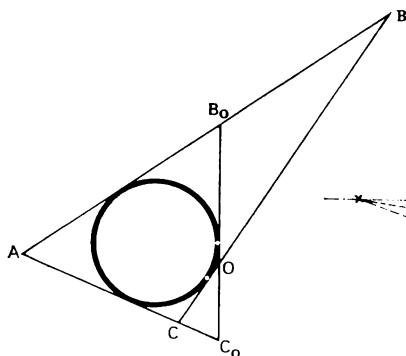


FIG. 67

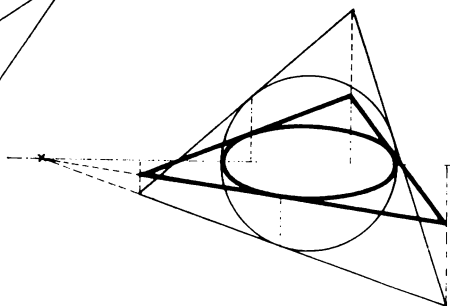


FIG. 68

construirse fácilmente (Fig. 68) a partir de la circunferencia mencionada.

Las «contracciones» del plano a una recta son sólo un caso particular de las transformaciones más generales, llamadas *afines*, del plano.

Transformaciones afines generales

Un par de vectores e_1, e_2 con un origen común O y no situados sobre una misma recta se llamará «sistema» de coordenadas del plano. Las coordenadas de un punto M del plano relativas a este referencial Oe_1e_2 serán entonces dos números (x, y) tales que, para alcanzar el punto M desde el origen O , es necesario superponer x veces a partir del punto O el vector e_1 , y luego y veces el vector e_2 . Este es un sistema de coordenadas cartesiano general en el plano. Análoga-

mente, se puede introducir un sistema de coordenadas cartesiano general en el espacio. El sistema más usual, llamado sistema de coordenadas cartesiano rectangular y que hemos venido utilizando ahora, corresponde al caso particular en que los vectores e_1, e_2 son mutuamente perpendiculares, y sus longitudes, iguales a la unidad de medida.

Una transformación afín general del plano es aquella en la cual una red dada de paralelogramos iguales se transforma en otra red de paralelogramos iguales. Expresado con mayor precisión: es una transformación del plano en la cual un sistema de coordenadas dado Oe_1e_2 se transforma en otro sistema (en general, con otra «métrica», es decir, con longitudes distintas para los vectores e'_1 y e'_2 y un ángulo diferente entre ellos), de suerte que un punto arbitrario M se transforma en otro punto M' que tiene las mismas coordenadas respecto al nuevo referencial que las de M respecto al antiguo (Fig. 69).

La «contracción» al eje Ox con coeficiente k es un caso especial en el que el referencial rectangular $Oe'_1e'_2$ se transforma en el $Oe'_1ke'_2$.

Puede demostrarse fácilmente que en una transformación afín toda recta se transforma en una recta, rectas paralelas en rectas paralelas y, si un punto divide a un segmento en una razón dada, su imagen divide a la imagen de este segmento en la misma razón. Además, puede demostrarse un teorema muy notable: cualquier transformación afín

del plano puede obtenerse realizando un cierto movimiento rígido del plano sobre sí mismo, y luego, en general, dos «contracciones» de

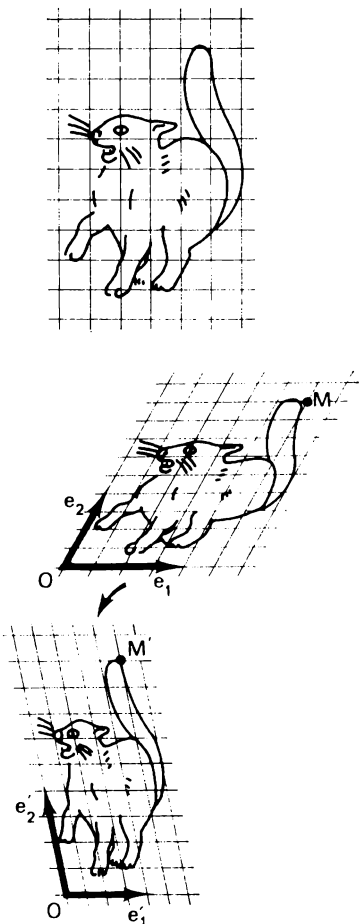


FIG. 69

coeficientes diferentes k_1 y k_2 hacia dos rectas mutuamente perpendiculares.

Para la demostración de este aserto, consideremos todos los radios de un círculo del plano (Fig. 70). Sea OA aquel radio que, después de la transformación, es el más corto, y apliquémoslo en $O'A'$. La perpendicular AB a OA se transforma entonces en $A'B'$, que debe ser perpendicular a $O'A'$, puesto que si la perpendicular $O'C'$ fuese

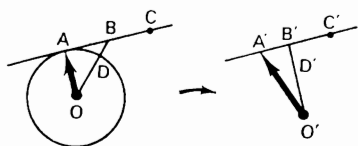


FIG. 70

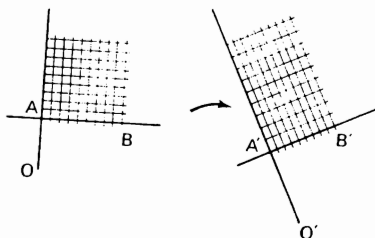


FIG. 71

diferente de $O'A'$, entonces sería imagen de la oblicua OC , y la imagen $O'D'$ del radio OD sería parte de la perpendicular $O'C'$, es decir, más corta que la oblicua $O'A'$, contrariamente a lo supuesto.

Las rectas mutuamente perpendiculares OA y AB se transforman por tanto en rectas mutuamente perpendiculares $O'A'$ y $A'B'$. Por tanto, la red cuadrada construida sobre OA y AB se transforma en una red de rectángulos iguales (Fig. 71); las «contracciones» uniformes se producen a lo largo de las rectas de esta red cuadrada.

De modo completamente análogo, una transformación afín general del espacio puede definirse como aquella en la cual el sistema espacial $Oe_1e_2e_3$ se transforma en el $O'e'_1e'_2e'_3$, con una «métrica» en general distinta, es decir: los segmentos unidad tendrán ahora una longitud distinta, los ángulos que forman entre sí también serán diferentes, y el punto M se transformará en el punto M' con las mismas coordenadas respecto al nuevo referencial que las del punto M respecto al antiguo.

Todas las propiedades enumeradas para el plano se verifican también para las transformaciones afines del espacio, excepto que en el último teorema será un movimiento rígido del espacio, seguido de tres «contracciones» hacia tres planos mutuamente perpendiculares, con ciertos coeficientes k_1, k_2, k_3 .

Aplicaciones de las transformaciones afines

Las aplicaciones más importantes de las transformaciones afines son:

1. En primer lugar su aplicación a la geometría para resolver problemas que se refieren a las propiedades afines de las figuras, es decir, propiedades que se conservan en las transformaciones afines. Ejemplos de ello son el teorema sobre los diámetros de una elipse y el de los triángulos circunscritos. Para resolver problemas de esta especie se hace una transformación afín de la figura en otra más sencilla, para la cual se demuestra la propiedad, y luego se vuelve a la figura original.

2. Segundo, su aplicación en geometría analítica para la clasificación de las curvas y superficies de segundo orden. El hecho importante es, como puede demostrarse, que las distintas elipses están relacionadas entre sí en el sentido de que una cualquiera puede obtenerse de otra mediante una transformación afín. Todas las hipérbolas son también afines entre sí, y lo mismo todas las parábolas. Pero mediante una transformación afín no se puede convertir una elipse en una parábola o una hipérbola en una parábola, es decir, no están relacionadas unas con otras por afinidad. Es natural dividir todas las curvas de segundo orden en clases afines (relacionadas por afinidad). La reducción de una ecuación a forma canónica proporciona exactamente esta clasificación, es decir, hay nueve clases afines de curvas de segundo orden. (No detallaremos por qué las elipses imaginarias y los pares de rectas paralelas imaginarias pertenecen a clases afines distintas. Realmente, ni en un caso ni en otro existe curva alguna en el plano. Lo que importa aquí son las propiedades algebraicas de la ecuación.)

De modo parecido, la clasificación de las superficies de segundo orden en 17 formas según sus ecuaciones canónicas es la misma que la clasificación afín.

Veamos un sencillito ejemplo de aplicación de la clasificación afín de las superficies de segundo orden. Demostraremos que si seleccionamos arbitrariamente en el espacio tres rectas a, b, c tales que: 1) dos cualesquiera de ellas no se encuentren en un mismo plano (es decir, que se crucen), y 2) no sean todas paralelas al mismo plano, entonces el conjunto d de todas las rectas del espacio que corten simultáneamente a las tres rectas a, b, c (Fig. 72) constituye la superficie de un hiperboloide de una hoja.

Examinemos más detenidamente este conjunto d de rectas. Por

un punto arbitrario A de la recta a podemos hacer pasar un plano P que contenga a la recta b y un plano Q que contenga a la recta c . Los planos P y Q se cortan en una única recta d , que pasa por el punto A de la recta a y corta a las rectas b y c . Dibujando todas estas rectas d que pasan por puntos de la recta a , obtenemos el conjunto de todas las rectas d del espacio que cortan a las tres rectas dadas a, b, c . Esta colección de rectas determina una superficie. Se com-

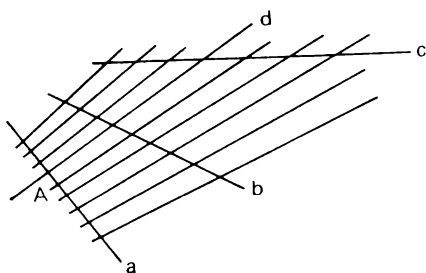


FIG. 72

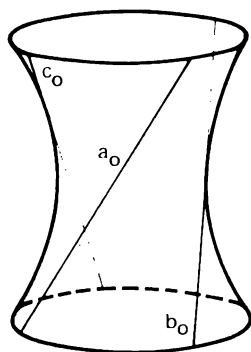


FIG. 73

prueba que cualquier hiperboloide de una hoja puede obtenerse de este modo, ya que basta tomar como recta a, b y c tres rectas distintas a_0, b_0, c_0 de una familia (Fig. 73), y como rectas d , todas las rectas de la otra familia. Recíprocamente, supongamos dadas tres rectas a, b y c en el espacio, de modo que se crucen dos a dos y no sean paralelas a un mismo plano. Puede demostrarse entonces que estas rectas formarán siempre las tres aristas (sin puntos comunes) de un paralelepípedo (Fig. 74). Después de construir tales paralelepípedos para las rectas dadas a, b, c y para tres rectas a_0, b_0, c_0 de una misma familia de un hiperboloide arbitrario de una hoja, efectuamos la transformación afín del espacio que transforma el paralelepípedo a_0, b_0, c_0 en el a, b, c ; evidentemente, esta transformación aplica el hiperboloide sobre la superficie en cuestión. Pero, de acuerdo con la clasificación afín de las superficies de segundo orden, la imagen afín de un hiperboloide de una hoja sigue siendo un hiperboloide de una hoja.

3. En tercer lugar está la aplicación a la teoría de las transformaciones continuas de los medios continuos, por ejemplo, en la teoría de la elasticidad, en la teoría de los campos eléctrico o mag-

nético, etc. Un elemento pequeño de un medio continuo se transforma «casi» afínmente. Se dice así que «localmente la transformación es lineal» (se llama lineal a una expresión de primer grado, y en la sección siguiente se verá que en geometría analítica las fórmulas que definen transformaciones afines son de primer grado). Esto resulta evidente en la figura 75. Sobre las líneas de la cuadrícula mayor la distorsión es claramente visible. Pero para trozos pequeños de una

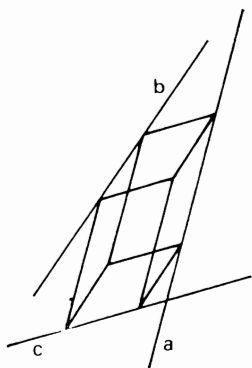


FIG. 74

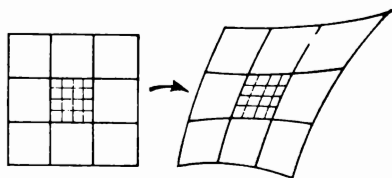


FIG. 75

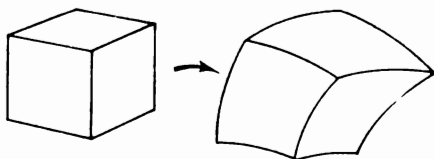


FIG. 76

red cuadrada muy densa, el efecto es muy pequeño, y la cuadrícula «casi» se transforma en una red de paralelogramos iguales. Algo semejante ocurre también en el espacio (Fig. 76). Del hecho de que toda transformación afín del espacio se reduce a un movimiento y a tres «contracciones» mutuamente perpendiculares, se sigue que bajo una deformación elástica cualquier elemento pequeño se mueve primero como un cuerpo rígido y luego experimenta tres «contracciones» mutuamente perpendiculares.

Fórmulas de las transformaciones afines

En una transformación afín en la que el referencial Oe_1e_2 se transforme en $O'e'_1e'_2$, siendo las coordenadas del nuevo origen O' relativas al antiguo ξ, η , y las coordenadas de los vectores e'_1 y e'_2 respecto al referencial antiguo a_1, a_2 y b_1, b_2 , respectivamente, las

fórmulas de la transformación afín, como puede deducirse fácilmente de la figura 77, son las siguientes:

$$x' = a_1x + b_1y + \xi,$$

$$y' = a_2x + b_2y + \eta$$

en el sentido de que si x, y , son las coordenadas de un punto M respecto al referencial antiguo Oe_1e_2 , los valores x', y' dados por estas fórmulas son las coordenadas, respecto al mismo referencial, de la imagen M' de este punto.

En efecto, sea $Oe'_1e'_2$ un referencial y $Oe'_1e'_2$ su imagen en una transformación afín, M un punto cualquiera del plano y M' su imagen. Entonces, por definición de transformación afín, si x, y son las coordenadas del punto M respecto al referencial Oe_1e_2 , las coordenadas de su imagen M' respecto a la imagen $Oe'_1e'_2$ de este referencial, son exactamente las mismas x, y .

Consideremos un vector $\mathbf{m}'$ que une el origen O del referencial antiguo a la imagen M' del punto M . Entonces $\mathbf{m}' = x'\mathbf{e}_1 + y'\mathbf{e}_2$. Pero este vector es igual a una cierta suma vectorial

$$\mathbf{m}' = \xi\mathbf{e}_1 + \eta\mathbf{e}_2 + x\mathbf{e}'_1 + y\mathbf{e}'_2$$

y los vectores $\mathbf{e}'_1$ y $\mathbf{e}'_2$ son

$$\mathbf{e}'_1 = a_1\mathbf{e}_1 + a_2\mathbf{e}_2, \quad \mathbf{e}'_2 = b_1\mathbf{e}_1 + b_2\mathbf{e}_2$$

por lo que

$$\mathbf{m}' = \xi\mathbf{e}_1 + \eta\mathbf{e}_2 + a_1x\mathbf{e}_1 + a_2x\mathbf{e}_2 + b_1y\mathbf{e}_1 + b_2y\mathbf{e}_2,$$

es decir,

$$\mathbf{m}' = (a_1x + b_1y + \xi)\mathbf{e}_1 + (a_2x + b_2y + \eta)\mathbf{e}_2.$$

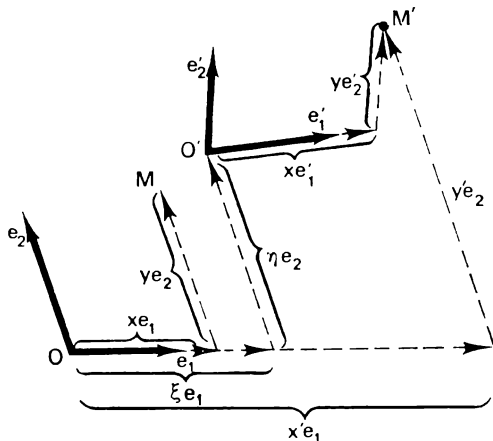


FIG. 77

Comparando esta expresión de m' con la primera obtenemos

$$\left. \begin{aligned} x' &= a_1x + b_1y + \xi, \\ y' &= a_2x + b_2y + \eta. \end{aligned} \right\} \quad (21)$$

El determinante

$$\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1,$$

como puede demostrarse, es distinto de cero e igual al cociente del área del paralelogramo construido sobre los vectores del nuevo referencial al área del mismo paralelogramo construido sobre los vectores del referencial antiguo.

Para el espacio se obtienen fórmulas análogas:

$$\left. \begin{aligned} x' &= a_1x + b_1y + c_1z + \xi, \\ y' &= a_2x + b_2y + c_2z + \eta, \\ z' &= a_3x + b_3y + c_3z + \zeta, \end{aligned} \right\} \quad (22)$$

donde (ξ, η, ζ) son las coordenadas del origen O' del referencial transformado $O'e'_1e'_2e'_3$ y (a_1, a_2, a_3) , (b_1, b_2, b_3) , (c_1, c_2, c_3) son las coordenadas de sus vectores e'_1, e'_2, e'_3 referidas al referencial antiguo $Oe_1e_2e_3$.

El determinante ⁵

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1$$

es distinto de cero e igual al cociente del volumen del paralelepípedo formado por los vectores del nuevo referencial al volumen del paralelepípedo formado por los vectores del antiguo.

Transformaciones ortogonales

Los movimientos rígidos del plano en sí mismo o dichos movimientos más una simetría respecto a una recta del plano, se llaman *transformaciones ortogonales del plano*, y los movimientos rígidos del

⁵ Sobre determinantes véase capítulo 15.

espacio, o estos movimientos más una simetría del espacio respecto a uno de sus planos, se llaman *transformaciones ortogonales del espacio*. Es claro que las transformaciones ortogonales son transformaciones afines en las cuales la «métrica» del referencial no cambia, puesto que sólo experimenta un movimiento rígido, o un movimiento más una reflexión.

Estudiaremos las transformaciones ortogonales en el caso de coordenadas rectangulares, es decir, cuando los vectores del referencial son perpendiculares entre sí y tienen longitudes iguales a la unidad de medida. Después de una transformación ortogonal los vectores del referencial siguen siendo mutuamente perpendiculares, es decir, su producto escalar es igual a cero, y sus longitudes, iguales a la unidad. Por tanto (véase fórmula (14) de este capítulo), en el caso del plano tenemos

$$a_1b_1 + a_2b_2 = 0, \quad a_1^2 + a_2^2 = 1, \quad b_1^2 + b_2^2 = 1, \quad (21')$$

y en el caso del espacio

$$\begin{aligned} a_1b_1 + a_2b_2 + a_3b_3 &= 0, & a_1^2 + a_2^2 + a_3^2 &= 1, \\ a_1c_1 + a_2c_2 + a_3c_3 &= 0, & b_1^2 + b_2^2 + b_3^2 &= 1, \\ b_1c_1 + b_2c_2 + b_3c_3 &= 0, & c_1^2 + c_2^2 + c_3^2 &= 1. \end{aligned} \quad (22')$$

Por tanto, si el referencial inicial se elige rectangular, las fórmulas (21) definen una transformación ortogonal si, y sólo si, se verifican las condiciones (21') de ortogonalidad; y las fórmulas (22) definen una transformación ortogonal del espacio si se satisfacen las condiciones (22') de ortogonalidad. Puede demostrarse que si $\Delta > 0$, tenemos un movimiento rígido, y si $\Delta < 0$, un movimiento rígido más una simetría.

§12. Teoría de los invariantes

El concepto de invariante. Invariantes de una ecuación de segundo grado en dos variables

En la segunda mitad del siglo pasado se introdujo un nuevo e importante concepto: el de invariante.

Consideremos, por ejemplo, un polinomio de segundo grado en dos variables

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F. \quad (23)$$

Si consideramos x, y , como coordenadas rectangulares y hacemos una transformación a nuevos ejes rectangulares, entonces, después de reemplazar x, y en (23) por sus expresiones en función de las nuevas coordenadas x', y' , de quitar paréntesis y de reducir términos semejantes, obtenemos un nuevo polinomio transformado con coeficientes distintos

$$A'x'^2 + 2B'x'y' + C'y'^2 + 2D'x' + 2E'y' + F'. \quad (24)$$

Pues bien, existen ciertas expresiones formadas a partir de los coeficientes cuyo valor numérico no cambia en la transformación, aunque varíen dichos coeficientes. Una expresión de este tipo en A', B', C', D', E' y F' tiene exactamente el mismo valor numérico que cuando se calcula con los A, B, C, D, E y F .

Las expresiones de este tipo se llaman invariantes del polinomio (23) respecto al grupo de transformaciones ortogonales (es decir, relativas a las transformaciones de un conjunto de coordenadas rectangulares x, y a cualesquiera otras coordenadas rectangulares x', y').

Invariantes de esta clase son:

$$I_1 = A + C,$$

$$I_2 = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = AC - B^2$$

$$I_3 = \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix} = ACF + 2BDE - AE^2 - CD^2 - FB^2,$$

es decir,

$$A + C = A' + C', \quad AC - B^2 = A'C' - B'^2,$$

$$\begin{aligned} ACF + 2BDE - AE^2 - CD^2 - FB^2 = \\ = A'C'F' + 2B'D'E' - A'E'^2 - C'D'^2 - F'B'^2. \end{aligned}$$

Se demuestra el importante teorema de que cualquier invariante ortogonal del polinomio (23) puede expresarse en función de estos tres invariantes básicos.

Igualando el polinomio (23) a cero, se obtiene la ecuación de una curva de segundo orden. Cualquier magnitud relacionada con esta curva, pero no con su posición en el plano, es independiente de

las coordenadas en que se exprese su ecuación; por tanto, escrita en función de los coeficientes será un invariante ortogonal del polinomio (23) y se podrá expresar en función de los tres invariantes básicos. Además, puesto que al multiplicar los seis coeficientes de la ecuación por cualquier número dado t (diferente de cero) la curva representada por la ecuación es la misma, la expresión de cualquier propiedad de la curva en función de I_1 , I_2 e I_3 debe ciertamente ser tal que si multiplicamos A, B, C, D, E, F por t , el número t no aparezca al simplificar. Se dice entonces que la ecuación es homogénea, de grado cero, respecto a A, B, C, D, E, F .

Comprobemos esto con un ejemplo. Supongamos que la ecuación

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0$$

representa una elipse. Puesto que la ecuación determina completamente la elipse, podemos calcular a partir de ella (es decir, de sus coeficientes) todas las magnitudes básicas relacionadas con la elipse. Por ejemplo, podemos calcular los semiejes a y b expresándolos en función de los coeficientes. Las expresiones de dichos semiejes serán invariantes y, por tanto, expresables en función de I_1, I_2, I_3 . Por reducción de la ecuación a la forma canónica y después de algunos cálculos se obtienen para los semiejes las siguientes expresiones, en función de I_1, I_2, I_3 :

$$\sqrt{\frac{2 |I_3|}{|I_2| |I_1 \pm \sqrt{I_1^2 - 4I_2}|}},$$

que son homogéneas respecto a A, B, C, D, E, F .

Se ve claramente que los invariantes I_1, I_2, I_3 , que son homogéneos pero no de grado cero, no tienen ningún significado geométrico; son entidades algebraicas.

También puede demostrarse que la expresión

$$K_1 = \begin{vmatrix} A & D \\ D & F \end{vmatrix} + \begin{vmatrix} C & E \\ E & F \end{vmatrix} = AF - D^2 + CF - E^2$$

varía en una traslación paralela, pero no por simple rotación de los ejes rectangulares dados, por lo cual se llama *semi-invariante*.

Como ejemplo de aplicación de los invariantes y semi-invariantes, damos la tabla 1, que, una vez calculados I_1, I_2, I_3 y K_1 , permite

determinar directamente a partir de su ecuación la clase afín de una curva de segundo orden.

En la tabla 1 se dan las condiciones necesarias y suficientes para que una ecuación de segundo orden sea reducible a una u otra de las nueve formas canónicas ($I_1 I_3$ designa el producto de I_1 e I_3).

TABLA 1

Criterio de la clase	Nombre	Ecuación de reducción	Ecuación canónica
$I_2 > 0, I_1 I_3 < 0$	elipse		$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$
$I_2 > 0, I_1 I_3 > 0$	elipse imaginaria		$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$
$I_2 > 0, I_3 = 0$	punto	$\lambda_1 x^2 + \lambda_2 y^2 + \frac{I_3}{I_2} = 0$	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$
$I_2 < 0, I_3 \neq 0$	hipérbola		$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$
$I_2 < 0, I_3 = 0$	par de rectas que se cortan		$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$
$I_2 = 0, I_3 \neq 0$	parábola	$I_1 x^2 + 2\sqrt{-\frac{I_3}{I_1}} y = 0$	$x^2 = 2py$
$I_2 = 0, I_3 = 0, K_1 < 0$	par de rectas paralelas		$x^2 = a^2$
$I_2 = 0, I_3 = 0, K_1 > 0$	par de rectas imaginarias paralelas	$I_1 x^2 + \frac{K_1}{I_1} = 0$	$x^2 = -a^2$
$I_2 = 0, I_3 = 0, K_1 = 0$	par de rectas coincidentes		$x^2 = 0$

Consideremos, por ejemplo, la ecuación $x^2 - 6xy + 5y^2 - 2x + 4y + 3 = 0$. Tenemos $A = 1$, $B = -3$, $C = 5$, $D = -1$, $E = 2$, $F = 3$, por lo que $I_1 = 6$, $I_2 = -4$, $I_3 = -9$. Puesto que se satisfacen las condiciones de la cuarta línea de la tabla: $I_2 < 0$, $I_3 \neq 0$, la ecuación es una hipérbola. Sus semiejes son

$$\sqrt{\frac{2 \cdot 9}{4 \cdot |6 \pm \sqrt{36 + 16}|}} \approx 0,57 \quad \text{y} \quad 1,93.$$

Los coeficientes de la ecuación reducida (I), (II) y (III) se expresan en función de los invariantes y semi-invariantes como sigue:

$$\lambda_1 x''^2 + \lambda_2 y''^2 + \frac{I_3}{I_2} = 0, \quad (\text{I})$$

$$I_1 x''^2 + 2 \sqrt{-\frac{I_3}{I_1}} y'' = 0, \quad (\text{II})$$

$$I_1 x''^2 + \frac{K_1}{I_1} = 0, \quad (\text{III})$$

donde λ_1 y λ_2 son las raíces de la llamada ecuación cuadrática característica:

$$\lambda^2 - I_1 \lambda + I_2 = 0.$$

Las fórmulas I-III permiten calcular rápidamente los semiejes a y b de una elipse y una hipérbola, el parámetro p de una parábola, y la distancia $2a$ entre rectas paralelas. Anteriormente dimos las fórmulas para los semiejes. El parámetro p es igual a

$$p = \sqrt{-\frac{I_3}{I_1^3}}, \text{ y la distancia } 2a = 2 \sqrt{-\frac{K_1}{I_1^2}}.$$

Análogamente podría darse una teoría de invariantes y semi-invariantes (con la correspondiente tabla para la determinación de la clase afín y las fórmulas de los coeficientes de las ecuaciones reducidas) para las superficies de segundo orden en el espacio tridimensional.

Hay que tener en cuenta que hasta ahora sólo nos hemos referido

a los invariantes que se utilizan en geometría analítica para curvas y superficies de segundo orden. El concepto de invariante, sin embargo, tiene un significado mucho más amplio.

Por invariante de un objeto, respecto a alguna de sus transformaciones, entendemos cualquier valor numérico, vectorial, etc., relacionado con este objeto, que no varíe en dicha transformación. En el problema anterior el objeto es un polinomio de segundo grado de dos variables (con mayor precisión, el conjunto de sus coeficientes), y las transformaciones son aquellas transformaciones del polinomio que se obtienen por el paso de un sistema de coordenadas rectangular a otro.

Otro ejemplo: el objeto es una masa dada de gas a temperatura también dada. Las transformaciones son los cambios de volumen o presión de esta masa de gas. El invariante, de acuerdo con la ley de Boyle-Mariotte, es el producto del volumen por la presión. Por otro lado, se puede hablar de las longitudes de segmentos en el espacio, o del valor de los ángulos, como invariantes del grupo de movimientos del espacio; o de la razón en la que un punto divide a un segmento, o de razones de áreas, como invariantes del grupo de las transformaciones afines del espacio, etc.

Muchos invariantes son particularmente importantes en la Física.

§13. Geometría proyectiva

Proyecciones perspectivas

Hace mucho tiempo que los artistas empezaron a estudiar las leyes de la perspectiva, lo cual era necesario porque el hombre ve los objetos tal y como se proyectan sobre la retina del ojo, de modo que la forma y localización de aquéllos quedan distorsionadas de una manera característica. Por ejemplo, los postes de telégrafo parecen más pequeños y próximos entre sí con la distancia, las vías paralelas del ferrocarril, parecen converger, etc. No consideraremos aquí perspectivas espaciales, es decir, las propiedades de las proyecciones perspectivas de los objetos del espacio sobre un plano, sino sólo las propiedades de las proyecciones perspectivas de un plano sobre otro plano.

Consideremos una fotografía P (por ejemplo, un fotograma de una película), una pantalla P' y, entre ambas, una lente S (Fig. 78). Entonces, si la fotografía es transparente y se ilumina por detrás (si no es transparente se iluminará por delante, es decir, del lado de

donde esté la lente), los puntos iluminados de la fotografía emitirán rayos de luz, que, una vez recogidos por la lente, aparecerán sobre la pantalla P' en forma de puntos. Supondremos que esta proyección tiene lugar como si los puntos de la fotografía P se proyectasen sobre la pantalla P' según rectas que pasaran por el centro óptico S de la lente.

La situación es muy sencilla cuando los planos P y P' son paralelos. En este caso se obtiene evidentemente sobre el plano P' una

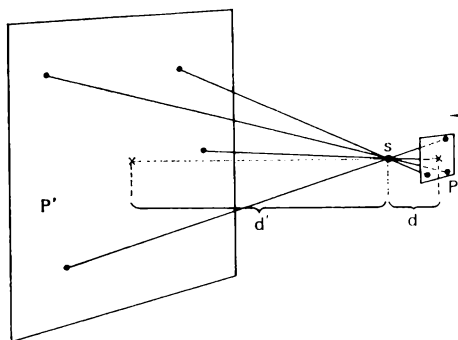


FIG. 78

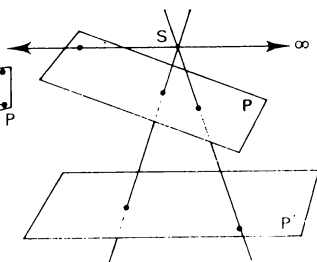


FIG. 79

imagen no distorsionada de todo lo que se encuentre en el plano P . Esta imagen será más pequeña o más grande que la original según que el cociente $d' : d$ (donde d y d' son las distancias desde el centro de la lente a los planos P y P' , respectivamente) sea mayor o menor que 1.

La situación es mucho más complicada cuando los planos P y P' no son paralelos (Fig. 79). En este caso, en una proyección a través del punto S no sólo varía el tamaño de la figura, sino que también se distorsiona su forma. Puede ocurrir que las líneas paralelas se conviertan en convergentes o que se modifique la razón en que un punto divide a un segmento, etc. En general, pueden variar algunas de las relaciones que permanecen invariantes en las transformaciones afines.

Esta clase de proyección se da, por ejemplo, en la fotografía aérea. El aeroplano oscila durante el vuelo, y por tanto el aparato fotográfico (Fig. 80a) sujeto rígidamente a él no está, en general,

colocado completamente en vertical; en el momento de la exposición generalmente se encuentra en posición oblicua, es decir, se obtiene una imagen distorsionada del terreno (que se supone plano).

¿Cómo corregir esta imagen? Para ello es necesario estudiar las propiedades de la proyección de un plano P sobre otro plano Π (en general, los dos planos no son paralelos) mediante rectas que pasan por un punto S que no pertenece ni al plano P ni al plano Π . Tales proyecciones se llaman *proyecciones perspectivas*.

Más tarde demostraremos el siguiente e importante teorema.

Teorema

Si tenemos dos proyecciones perspectivas de un plano P sobre un plano Π tal que en ambas proyecciones los cuatro puntos A, B, C, D de «posición general» sobre el plano P (es decir, cuatro puntos tales que no haya tres alineados) se proyectan en los mismos puntos A', B', C', D' del plano Π , entonces todos los puntos del plano P también se proyectarán, en ambas proyecciones, sobre los mismos puntos del plano Π .

En otras palabras, una proyección perspectiva queda completamente determinada si se conoce la proyección de cuatro puntos arbitrarios «de posición general» de la figura a proyectar.

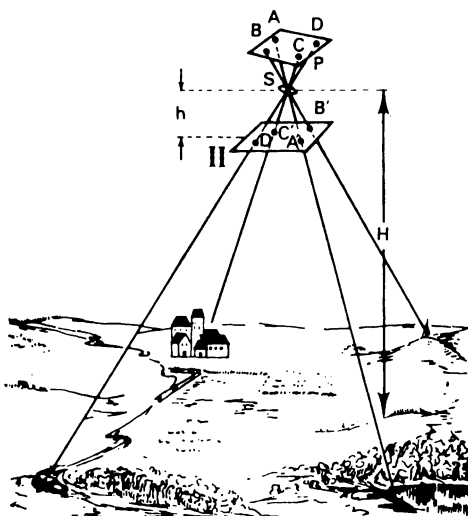
Este es el llamado teorema fundamental de la perspectividad plana.

Aplicación del teorema fundamental de la perspectividad plana a la fotografía aérea

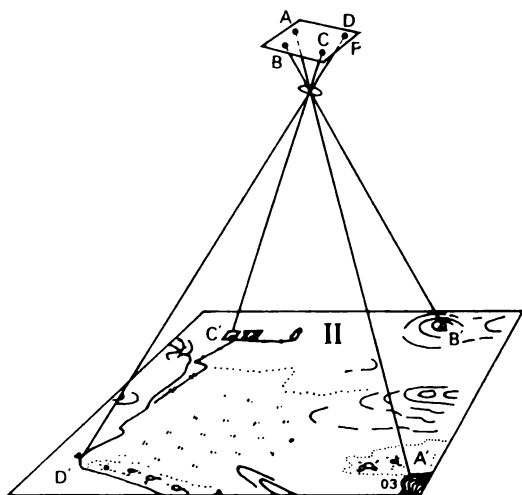
Vamos a ver cómo este teorema proporciona un método conveniente para corregir la imagen fotográfica distorsionada.

Si imaginamos que en el momento de la exposición aérea colocamos una pantalla horizontal Π situada a la distancia h bajo el centro S de la lente (Fig. 80a), entonces la proyección (sobre esta pantalla y a través del centro S) de la imagen recogida sobre la placa fotográfica P no estará distorsionada, sino que será semejante a la región fotografiada con una escala $h : H$, donde H es la altura del aeroplano en el momento de la exposición. Para corregir la imagen fotográfica P y convertirla en una imagen no distorsionada, se procede del modo siguiente. La fotografía revelada P se sitúa en un aparato de proyección colocado sobre un trípode especial sobre el cual, por medio de tornillos ajustables, puede acercarse o alejarse el aparato y ser enfocado en cualquier dirección.

Fijamos ahora a la pantalla Π (Fig. 80b) un mapa topográfico



(a)



(b)

FIG. 80

de la localidad, trazado con ayuda de medidas efectuadas sobre la superficie de la Tierra (no un mapa detallado, puesto que los detalles de interés los proporciona precisamente la fotografía aérea). Sobre este mapa sujeto a la pantalla II se seleccionan cuatro puntos, A' , B' , C' , D' que puedan localizarse fácilmente también sobre la fotografía (por ejemplo, un cruce de carreteras, la esquina de una casa, etc.), y en los puntos correspondientes A , B , C , D de la fotografía P perforamos la película con una aguja. Situamos entonces una lámpara de proyección detrás de la placa P en posición tal que la imagen se proyecte sobre la pantalla II a través de una lente S del aparato soporte. Utilizando los tornillos de ajuste hacemos que los rayos de luz que atraviesan los cuatro agujeros caigan sobre los puntos correspondientes A' , B' , C' , D' del mapa sujeto a la pantalla. Después se reemplaza el mapa topográfico por una placa fotográfica y sin modificar la posición de los tornillos fotografiamos la imagen proyectada sobre la pantalla II de la imagen P tomada por el aeroplano.

Según el teorema anteriormente enunciado, se obtiene así un mapa verdadero (es decir, semejante al de la localidad), no distorsionado, de la región fotografiada.

Pasemos ahora a presentar la teoría necesaria para la demostración del teorema fundamental.

El plano proyectivo

El conjunto de todas las rectas y planos del espacio que pasan por un punto dado S del espacio se llama haz proyectante de rectas y planos de centro S . Si se corta este haz por un plano P que no pase por el centro, a todo punto del plano P le corresponderá una recta del haz que corta al plano P en este punto, y a cada recta del plano P le corresponderá el plano del haz que corta a P según esta recta. Sin embargo, no se establece así una aplicación uno-a-uno del conjunto de las rectas y planos del haz en el conjunto de los puntos y rectas del plano P . En efecto, a las rectas y planos del haz que son paralelos al plano P no les corresponde en esta relación ningún punto o recta del plano P , puesto que no lo cortan. Sin embargo, *diremos* que estas rectas del haz cortan al plano P en sus puntos ideales (o infinitamente distantes), situados en las direcciones correspondientes, y que un tal plano del haz corta al plano P según una recta ideal (o infinitamente distante). El plano P , completado con estos puntos y rectas ideales, se llama *plano completado o proyectivo* y lo notaremos por P^* . El conjunto de rectas y planos del haz S se aplica así uno-a-uno sobre

el conjunto de los puntos y rectas (reales e ideales) del plano proyectivo P^* .

Por tanto, *diremos* que un punto (real o ideal) pertenece a una recta (real o ideal) del plano proyectivo P^* si la recta correspondiente del haz se encuentra en el correspondiente plano del haz. Desde este punto de vista, dos rectas cualesquiera del plano proyectivo se cortan (en un punto real o ideal), puesto que dos planos cualesquiera del haz se cortan según una recta del haz. De aquí se deduce, entre otras cosas, que la recta ideal está constituida, sencillamente, por el conjunto de todos los puntos ideales.

En esencia, la completación del plano con los elementos ideales significa que se utiliza este plano como sección transversal para estudiar el haz de todas las rectas y planos que pasan por un punto.

Aplicaciones proyectivas; teorema fundamental

Por *aplicación proyectiva* entendemos una aplicación de un plano proyectivo P^* sobre otro plano proyectivo $P^{*'}$ (que puede coincidir con el plano P^* , y en ese caso se habla de transformación proyectiva del plano P^*) que sea, en primer lugar, una aplicación puntual uno-a-uno, y en segundo lugar, una aplicación tal que los conjuntos colineales de puntos del plano P^* se apliquen en conjuntos colineales de puntos del plano $P^{*'}$, y recíprocamente. (Por puntos y rectas entenderemos siempre tanto los puntos y rectas reales como los ideales.)

Es claro que dos proyecciones perspectivas arbitrarias del plano P^* sobre un plano Π^* se pueden obtener la una de la otra mediante transformaciones proyectivas.

En efecto, 1) sus puntos (reales o ideales) están en correspondencia uno-a-uno con los puntos (reales o ideales) del plano proyectivo P^* , y 2) a puntos colineales de la primera proyección corresponden puntos colineales del plano P^* , y por tanto también de la segunda proyección, y recíprocamente. Así, pues, el teorema enunciado de la teoría de las perspectividades es una consecuencia directa del siguiente teorema sobre transformaciones proyectivas: si en una transformación proyectiva del plano Π^* , cuatro de sus puntos A, B, C, D , no alineados tres a tres, permanecen fijos, todos sus puntos permanecen fijos.

A continuación bosquejaremos la idea de la demostración de este teorema por medio de la llamada red de Möbius.

Observemos que: 1) si en una transformación proyectiva dos puntos permanecen fijos, la recta que pasa por ellos se transforma en sí misma, y 2) si dos rectas se aplican en sí mismas, el punto de su

intersección permanece fijo. Por tanto, del hecho de que los puntos A, B, C, D del plano Π^* permanezcan fijos, se sigue que también los puntos E, F, G, H, K, L , etc., permanecerán fijos (Fig. 81). Puede continuarse la construcción de tales puntos uniendo los ya obtenidos. Esta es la llamada red de Möbius. Continuando su construcción, podemos situar los puntos tan próximos como deseemos. Puede demostrarse que el conjunto de estos puntos cubre densamente todo

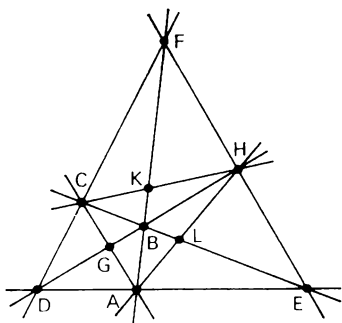


FIG. 81

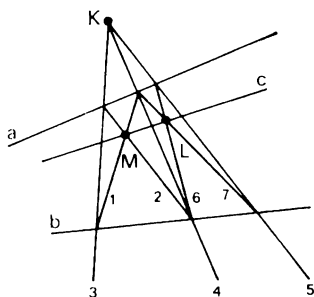


FIG. 82

el plano. Por tanto, suponiendo la continuidad de la transformación proyectiva (que, de hecho, se deduce de su definición, aunque la demostración no es fácil), el resultado es que, si en una transformación proyectiva del plano Π^* los puntos A, B, C, D permanecen fijos, todos los puntos del plano Π^* permanecerán fijos.

Geometría proyectiva

Por *geometría proyectiva bidimensional* entendemos el conjunto de todos los teoremas referentes a aquellas propiedades de las figuras del plano proyectivo —es decir, el plano ordinario completado con los elementos ideales— que no se modifican en las transformaciones proyectivas.

He aquí un ejemplo de un problema de geometría proyectiva. Dadas dos rectas a y b y un punto M (Fig. 82), trazar la recta c que pase por el punto M y por el punto de intersección de las rectas a y b , sin utilizar este punto de intersección (como puede ocurrir, si este punto está muy distante). Si por el punto M dibujamos las dos

secantes 1 y 2 y luego las rectas 3 y 4 por los puntos de intersección con las rectas a y b , obtenemos el punto K . Dibujamos la recta 5 que pasa por él y las secantes 6 y 7; puede demostrarse entonces que la recta c , que pasa por el punto L de intersección de las rectas 6 y 7 y el punto M , es la recta buscada.

De la teoría de las secciones cónicas se sigue (Fig. 83) que la elipse, hipérbola y parábola son proyecciones perspectivas unas de

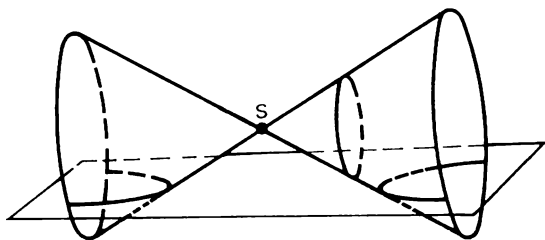


FIG. 83

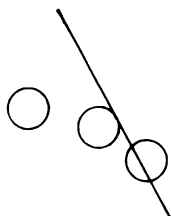


FIG. 84

las otras, y que todas ellas son proyecciones perspectivas de la circunferencia.

Si consideramos las proyecciones perspectivas como transformaciones proyectivas de los planos proyectivos P^* y P^{**} uno sobre el otro, por superposición de estos planos llegamos al resultado de que todas las elipses, hipérbolas y parábolas son transformaciones proyectivas de la circunferencia. La diferencia entre ellas es que las imágenes proyectivas de la circunferencia en las transformaciones en las que una recta que no corta al círculo se transforma en una recta infinitamente distante, son elipses; por otra parte, si una recta tangente al círculo se transforma en una recta infinitamente distante, se obtiene una parábola, y si en una secante, una hipérbola (Fig. 84).

Fórmulas de las transformaciones proyectivas

Si el plano P^* se define un sistema de coordenadas cartesianas ordinarias, entonces, como puede demostrarse, las fórmulas para las transformaciones proyectivas del plano son las siguientes:

$$x' = \frac{a_1x + b_1y + c_1}{a_3x + b_3y + c_3}, \quad y' = \frac{a_2x + b_2y + c_2}{a_3x + b_3y + c_3},$$

donde el determinante

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \neq 0,$$

y recíprocamente.

Si para algún punto (x, y) los denominadores se anulan, esto significa que su imagen (x', y') es un punto ideal (infinitamente distante). La ecuación

$$a_3x + b_3y + c_3 = 0$$

representa la recta que en la transformación proyectiva dada tiene por imagen la recta ideal (infinitamente distante).

§14. Transformaciones de Lorentz

Deducción de las fórmulas de la transformación de Lorentz para los movimientos sobre una línea recta y sobre el plano, a partir de la condición de constancia de la velocidad de la luz

Al finalizar el siglo XIX se descubrió una contradicción fundamental en física. El conocido experimento de Michelson —en el que se midió la velocidad de la luz (que es aproximadamente de 300.000 km/seg.) tanto en la dirección del movimiento de la Tierra a lo largo de su órbita alrededor del Sol (la velocidad de la Tierra es de unos 30 km/seg.) como a lo largo de la perpendicular a esa dirección— demostró irrefutablemente que todos los cuerpos que se mueven en la naturaleza, incluso si se mueven en el vacío, se contraen en la dirección del movimiento. La teoría de esta contracción fue investigada detalladamente por el físico Lorentz. Demostró que dicha contracción es mayor cuanto más próxima a la velocidad de la luz en el vacío sea la velocidad del cuerpo en movimiento, y que a una velocidad igual a la de la luz la contracción sería infinita. Lorentz dedujo las fórmulas para esta contracción. Pero muy poco después el físico Einstein introdujo en este problema un punto de vista completamente distinto, al que ya se había acercado Poincaré. Einstein razonó como sigue. Si suponemos que para la propagación de la luz, como para el movimiento ordinario de un cuerpo material, es válida la ley de composición de velocidades de Galileo, entonces la velocidad de la luz será $c' = c + v$, donde v es la velocidad del obser-

vador que se mueve hacia el manantial de luz, y c es la velocidad de la luz para un observador en reposo. Del experimento de Michelson se deduce que $c' = c$. La ley $c' = c + v$ se basa en la transformación

$$\begin{aligned}x' &= x + v_x t, \\t' &= t,\end{aligned}\tag{25}$$

que relaciona la coordenada x de un punto respecto a un sistema de coordenadas I, con su coordenada x' respecto a un sistema de coordenadas II, que tiene sus ejes paralelos a los del sistema I y se mueve paralelamente al eje Ox con velocidad v_x , respecto al sistema I. Evidentemente, estas son las fórmulas que, según Einstein, deben modificarse.

Puede demostrarse (como ha hecho recientemente, por ejemplo, A. D. Alexandrov) que de la igualdad de la velocidad de la luz en ambos sistemas de coordenadas x, y, z, t y x', y', z', t' se deduce ya que las fórmulas de transformación de las coordenadas x, y, z, t a las coordenadas x', y', z', t' son lineales y homogéneas, es decir, tienen la forma

$$\begin{aligned}x' &= a_1x + b_1y + c_1z + d_1t, \\y' &= a_2x + b_2y + c_2z + d_2t, \\z' &= a_3x + b_3y + c_3z + d_3t, \\t' &= a_4x + b_4y + c_4z + d_4t.\end{aligned}\tag{26}$$

Por otras consideraciones se puede demostrar que su determinante⁶ es igual a la unidad.

Si un punto del sistema I se mueve rectilínea y uniformemente en una dirección arbitraria dada, con la velocidad de la luz c , entonces $x = v_x t$, $y = v_y t$, $z = v_z t$ y $v_x^2 + v_y^2 + v_z^2 = c^2$, de donde

$$x^2 + y^2 + z^2 - c^2 t^2 = 0.\tag{27}$$

Pero del experimento de Michelson se deduce que en el sistema II este punto necesariamente se mueve también con la misma velocidad de la luz c , para lo cual es preciso que

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 = 0.$$

⁶ Véase capítulo 15.

Por tanto, las fórmulas (26) no son simples transformaciones arbitrarias lineales, homogéneas y de determinante igual a 1, sino que además deben satisfacer la condición de que si las coordenadas x, y, z, t , son tales que

$$x^2 + y^2 + z^2 - c^2 t^2 = 0,$$

entonces las coordenadas transformadas x', y', z', t' deben satisfacer también esta ecuación. Tales transformaciones (26) se llaman *transformaciones de Lorentz*.

Consideremos en primer lugar el caso más sencillo: el de un punto que se mueve a lo largo del eje Ox . En este caso las fórmulas (26) son de la forma

$$x' = a_1 x + d_1 t, \quad (26')$$

$$t' = a_2 x + d_2 t,$$

y la ecuación (27)

$$x^2 - c^2 t^2 = 0. \quad (27')$$

Introduciendo la notación $ct = u$, las fórmulas (26') y la ecuación (27') toman la forma

$$x' = a_1 x + \frac{d_1}{c} u, \quad (26_1)$$

$$u' = a_2 c x + \frac{d_2 c}{c} u$$

y

$$x^2 - u^2 = 0. \quad (27_1)$$

Busquemos ahora la forma explícita de las fórmulas (26₁). Consideremos x y u como un sistema rectangular cartesiano en el plano, es decir, consideremos el problema geoméricamente; podemos entonces considerar las fórmulas (26₁) como las de una transformación afín del plano Oxu (cuyo determinante, como se demostró, es igual a 1). Notemos esta transformación por L . Si, como suponemos, $x^2 - u^2 = 0$ implica $x'^2 - u'^2 = 0$, entonces esta transformación aplica las rectas que se cortan

$$x^2 - u^2 = 0$$

sobre ellas mismas. La transformación L es por tanto una combinación de una contracción y una dilatación con coeficientes idénticos τ a lo largo de estas líneas.

De la figura 85 obtenemos:

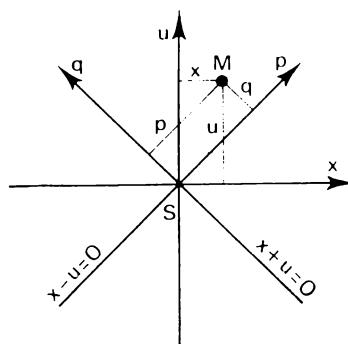


FIG. 85

$$x = \frac{p}{\sqrt{2}} - \frac{q}{\sqrt{2}},$$

$$u = \frac{p}{\sqrt{2}} + \frac{q}{\sqrt{2}}.$$

Pero después de la transformación L , los números p y q se transformarán en $p' = p/\tau$ y $q' = q\tau$, por lo que

$$x' \sqrt{2} = \frac{p}{\tau} - q\tau,$$

$$u' \sqrt{2} = \frac{p}{\tau} + q\tau.$$

Expresando p y q en función de x y u en el primer par de ecuaciones, sustituyendo en la segunda y simplificando, obtenemos

$$x' = \frac{x - \frac{\tau^2 - 1}{\tau^2 + 1} ct}{\frac{2\tau}{\tau^2 + 1}}; \quad t' = \frac{t - \frac{1}{c} \frac{\tau^2 - 1}{\tau^2 + 1} x}{\frac{2\tau}{\tau^2 + 1}},$$

o bien, poniendo $[(\tau^2 - 1)/(\tau^2 + 1)]c = v$, tenemos

$$x' = \frac{x - vt}{\sqrt{1 - (v/c)^2}}, \quad t' = \frac{t - (vx/c^2)}{\sqrt{1 - (v/c)^2}},$$

que son las conocidas *fórmulas de Lorentz*.

En particular, si tomamos $x = 0$, es decir, si consideramos el movimiento del origen del sistema de coordenadas I , obtenemos

$$x' = \frac{-vt}{\sqrt{1 - (v/c)^2}}, \quad t' = \frac{t}{\sqrt{1 - (v/c)^2}},$$

o bien $x' = -vt'$, por lo que evidentemente v es la velocidad del movimiento del sistema de coordenadas II respecto al sistema I.

Supongamos, por ejemplo, dos puntos sobre el eje Ox , de coordenadas x_1 y x_2 respecto al sistema I, por lo que la distancia entre ellos en el sistema I será $r = |x_1 - x_2|$. Veamos cuál es dicha distancia para un observador situado en el sistema II. Tenemos

$$x'_1 = \frac{x_1 - vt}{\sqrt{1 - (v/c)^2}}, \quad x'_2 = \frac{x_2 - vt}{\sqrt{1 - (v/c)^2}},$$

de donde

$$r' = |x'_1 - x'_2| = \frac{|x_1 - x_2|}{\sqrt{1 - (v/c)^2}}.$$

El factor $\sqrt{1 - (v/c)^2}$ es exactamente el coeficiente de la contracción de Lorentz. Puesto que c es muy grande, para v no muy grande este coeficiente es aproximadamente igual a 1, y por tanto la contracción no es significativa. Pero las partículas elementales, tales como los electrones o los positrones, se mueven frecuentemente con velocidades comparables a la de la luz, y por tanto, para estudiar su movimiento, es necesario tener en cuenta esta contracción o, como se suele decir, considerar el efecto relativista.

Pasemos ahora a un caso más complejo, aquel en el que el punto se mueve en el plano Oxy . Para este caso la transformación (26) tendrá la forma

$$\begin{aligned} x' &= a_1x + b_1y + d_1t, \\ y' &= a_2x + b_2y + d_2t, \\ t' &= a_3x + b_3y + d_3t, \end{aligned} \tag{26''}$$

donde

$$\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix} = 1,$$

y la ecuación (27) será

$$x^2 + y^2 - c^2t^2 = 0. \tag{27''}$$

Estas son las fórmulas de Lorentz para el movimiento en el plano Oxy .

Pongamos otra vez $ct = u$. Entonces la transformación (26'') puede escribirse como

$$\begin{aligned}x' &= a_1 x + b_1 y + \frac{d_1}{c} u, \\y' &= a_2 x + b_2 y + \frac{d_2}{c} u, \\u' &= a_3 cx + b_3 cy + \frac{d_3 c}{c} u,\end{aligned}\tag{26_2}$$

donde el valor del determinante seguirá siendo igual a 1, y la ecuación (27'') tomará la forma más sencilla

$$x^2 + y^2 - u^2 = 0.\tag{27_2}$$

Consideraremos x, y, u como las coordenadas rectangulares cartesianas de un punto del espacio tridimensional ordinario, y las fórmulas (26₂), como las de una transformación afín de este espacio. La ecuación (27₂) representa un cono circular recto K con un ángulo de 90° en el vértice (Fig. 86).

Desde el punto de vista de esta interpretación geométrica (la llamamos geométrica porque aquí consideraremos $u = ct$ simplemente como una coordenada espacial) de la transformación de Lorentz, el conjunto de los movimientos del plano es idéntico al conjunto de todas las transformaciones equi-afines del espacio (es decir, afines y conservadoras del volumen) que aplican el cono K sobre sí mismo.

Consideremos algunas transformaciones de Lorentz especiales.

1. Es claro que cualquier rotación rígida ω sobre el eje del cono K es una transformación equi-afín del espacio que aplica el cono K en sí mismo, es decir, es una transformación de Lorentz especial. La notaremos por ω .

2. Las simetrías especulares cuyo plano π pase por el eje del cono K son también transformaciones de Lorentz. Las notaremos por π .

3. Finalmente, consideremos la siguiente transformación (Fig. 87). Sean v y w un par cualquiera de generatrices opuestas del cono, y sean P y Q planos tangentes al cono a lo largo de estas generatrices.

Estos planos son mutuamente perpendiculares. Hagamos una contracción del espacio hacia el plano P y una dilatación, con el mismo coeficiente, desde el plano Q . Tal transformación del espacio es también afín y conserva todos los volúmenes. La notaremos por L . Demostremos que esta transformación aplica el cono K en sí mismo. Como el cono K tiene al eje u como eje de revolución, cualquier figura puede girarse de tal modo que las generatrices v y w queden situadas, por ejemplo, en el plano Sxu . Por tanto, es suficiente hacer la demostración para este caso.

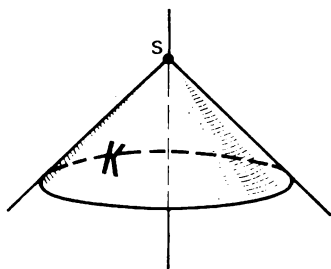


FIG. 86

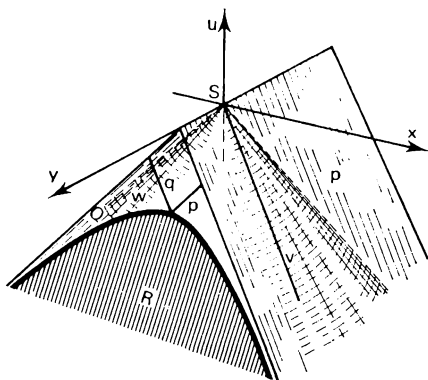


FIG. 87

Para ello cortamos al cono K por un plano arbitrario R paralelo al Sxu . La ecuación de este plano es $y = b$, donde b es constante. Sustituyendo este valor en la ecuación del cono K , obtenemos

$$x^2 - u^2 = -b^2.$$

Esta es la ecuación de una hipérbola en la que las rectas de intersección del plano R con los planos P y Q son justamente las asíntotas. Pero, puesto que para un punto de esa hipérbola el producto de sus distancias p y q a las asíntotas, es decir, a los planos P y Q , es constante, en la transformación L todos los puntos de esta hipérbola permanecen sobre la misma hipérbola, que por tanto se aplica sobre

sí misma. Ahora bien, toda la superficie del cono K consta de tales hipérbolas, y por tanto en la transformación L del espacio el cono K se transforma en sí mismo. La transformación L es, por consiguiente, también una transformación de Lorentz.

Puesto que en las transformaciones afines las rectas se transforman en rectas y las líneas que se cortan en líneas que se cortan, en una transformación de Lorentz cualquier haz S de rectas se aplica uno-a-uno sobre sí mismo. Además, en las transformaciones afines del espacio todos los planos se transforman en planos, de modo que tales transformaciones del haz en sí mismo son transformaciones proyectivas del haz. Si cortamos este haz por un plano Π perpendicular al eje del cono K (que se aplica en sí mismo en la transformación de Lorentz dada del espacio), extendemos dicho plano al plano proyectivo Π^* y señalamos los puntos de intersección de las rectas del haz S con el plano Π^* , obtenemos el resultado de que las transformaciones de Lorentz del haz producirán simultáneamente transformaciones proyectivas A del plano Π^* ; y estas últimas transformarán la circunferencia α , según la cual el plano Π^* corta a la parte interior del cono K , en sí misma. Para analizar las propiedades de las transformaciones de Lorentz es más fácil examinar tales transformaciones proyectivas A de la circunferencia α en sí misma.

Transformaciones proyectivas de una circunferencia en sí misma

Un punto, un rayo o semirrecta que parta de él, y uno de los semiplanos limitados por la recta se llamará un «referencial» del plano Π^* (no debe ser confundido con un referencial de coordenadas, §11). Demostraremos (Fig. 88) que si tomamos dos referenciales arbitrarios M y M' que contengan puntos interiores de la circunferencia α , entonces por medio de las transformaciones L , ω , π , podemos aplicar uno de estos referenciales en el otro. Para ello es suficiente hacer la transformación $A = L_1 \cdot \omega \cdot L_2^{-1}$ (o bien $A = L_1 \cdot \omega \cdot \pi \cdot L_2^{-1}$). La transformación L_1 lleva el primer referencial M al centro O de la circunferencia α ; la transformación ω le imprime el giro necesario, y finalmente la transformación L_2^{-1} lo hace coincidir con el segundo referencial M' .

Veamos, además, que hay una sola transformación A que aplica un referencial dado M en otro referencial dado M' . Para ello observemos primero que si hubiese dos transformaciones A_1 y A_2 que transformasen el referencial M en el M' , la transformación $A = A_1 A_2^{-1}$ no sería la transformación idéntica A y aplicaría al referencial M en

sí mismo. Por tanto, es suficiente demostrar que, si una transformación A aplica el referencial M en sí mismo, entonces es la identidad, es decir, deja fijos todos los puntos del plano de la circunferencia α .

Demostremoslo. Supongamos que la transformación A aplica el referencial M en sí mismo (Fig. 89). Entonces la recta AB de este referencial se aplica en sí misma; pero, puesto que transforma la circunferencia α en sí misma, o bien deja los puntos A y B fijos, o los intercambia. Esto último, sin embargo, es imposible, toda vez que la semirrecta del referencial se aplica en sí misma. Tracemos las tangentes a la circunferencia

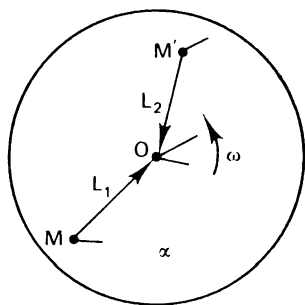


FIG. 88

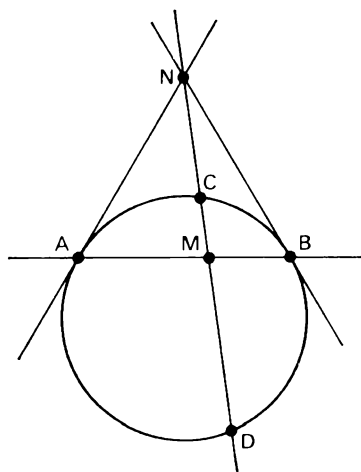


FIG. 89

α en los puntos A y B . Estas se transforman en sí mismas, puesto que, si una tangente se transformase en una secante $A\bar{A}$, entonces la transformación inversa aplicaría los puntos distintos A y $\bar{A}$ de la circunferencia α en el único punto A . Pero las A son transformaciones proyectivas y, por consiguiente, aplicaciones uno-a-uno. Puesto que en las transformaciones A estas tangentes se aplican en sí mismas, el punto M de su intersección permanece fijo, y por tanto la recta MN se aplica en sí misma. Del hecho de que la semirrecta del referencial M se aplica en sí misma concluimos como antes que los puntos C y D no se intercambian, sino que permanecen fijos. Por consiguiente, en la transformación proyectiva A del plano proyectivo Π^* , cuatro de sus puntos, A, B, C, D , no alineados tres a tres, permanecen fijos. Según el teorema de unicidad de las transformaciones proyectivas esta transformación es la identidad.

En el §5 del capítulo 16 se demostrará que utilizando las propiedades del grupo de Lorentz es fácil construir un modelo de la geometría plana de Lobachevski; y, si consideramos transformaciones de Lorentz para el caso general del movimiento de un punto en el espacio, entonces podemos hacer lo mismo para la geometría del espacio de Lobachevski, y probar así su coherencia.

Vemos que la teoría de las transformaciones de Lorentz, la geometría proyectiva y la teoría de la perspectividad, y la geometría no euclídea están claramente relacionadas las unas con la otra. Se verá que aún hay otra teoría que está también íntimamente relacionada con ellas: la de las transformaciones conformes en la teoría de funciones de una variable compleja, que resuelve problemas de la física matemática tan importantes como la distribución de temperaturas en una placa, el flujo de aire alrededor de las alas de un aeroplano, la distribución de carga en un campo electrostático plano, los problemas de la elasticidad en el plano, y muchos otros.

Conclusión

La geometría analítica es un método absolutamente indispensable para la investigación en otras ramas de la matemática, de la física y otras ciencias de la naturaleza. Por tanto, no sólo se estudia en las universidades, sino en todos los centros técnicos superiores, así como en algunas escuelas profesionales. Sería cuestión de pensar si no debería incluirse un tratamiento más o menos detallado de los elementos de la geometría analítica en los cursos superiores de bachillerato.

Distintos tipos de coordenadas

Los elementos esenciales de la geometría analítica son, como hemos visto, el método de las coordenadas y el estudio de las ecuaciones que relacionan entre sí estas coordenadas. Además de las coordenadas cartesianas existen otras. Por ejemplo, en el plano podemos elegir un punto P (llamado polo) y un rayo que parta de él (eje polar) y determinar la posición de un punto M por la longitud ρ del radio polar desde el polo al punto y por el valor ω del ángulo formado por este radio con el eje polar (Fig. 90).

En particular, tomando como polo uno de los focos de una elipse, hipérbola o parábola, y por eje polar el rayo que parte del foco según

el eje de simetría hacia el lado opuesto del vértice más próximo (Fig. 91), las tres figuras mencionadas tienen la misma ecuación

$$\varrho = \frac{p}{1 - \epsilon \cos \omega},$$

donde ϵ es la excentricidad de la curva, y p , su parámetro. Esta ecuación es de gran importancia en astronomía. Con su ayuda se demostró, a partir de las leyes de la inercia y de la gravitación universal, el resultado de que los planetas giran alrededor del Sol según elipses.

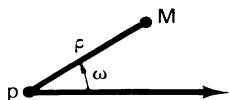


FIG. 90

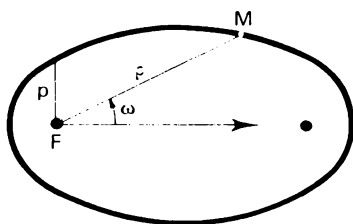


FIG. 91

También son conocidas las coordenadas geográficas —longitud y latitud— que determinan la posición de un punto sobre una esfera.

Análogamente, podemos tomar una malla de coordenadas sobre una superficie arbitraria, como se hace en geometría diferencial (ver capítulo 7), etc.

Geometría analítica n-dimensional e infinito-dimensional; geometría algebraica

Parecía que la geometría analítica, al haber experimentado un desarrollo tan enorme en el siglo XIX —descrito antes de un modo general— y haber producido tantas nuevas ideas, tendría necesariamente que haberse agotado; pero no ha ocurrido así. En tiempos muy recientes, dos nuevas y amplias ramas de la matemática se han desarrollado rápidamente y han venido a ampliar los conceptos de la geometría analítica: el *análisis funcional* y la *geometría algebraica*. Es cierto que estas dos ramas sólo representan a medias una continuación de la geometría analítica clásica: casi todo el análisis funcional es análisis, y gran parte de la geometría algebraica es teoría de funciones y topología.

Aclaremos esto último. A mediados del siglo pasado ya habían empezado los matemáticos a ocuparse de la geometría analítica cuatridimensional y en general n -dimensional, es decir, a estudiar aquellas cuestiones del álgebra que son generalizaciones directas de las cuestiones algebraicas del tipo que aparece en la geometría analítica bi- y tri-dimensional, al caso de cuatro o n variables. A finales del siglo XIX una serie de relevantes matemáticos tuvo la idea de que

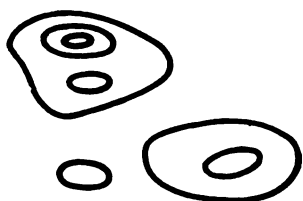


FIG. 92

para el análisis y la física matemática era importante considerar la geometría analítica de infinitas dimensiones.

En principio cabría pensar que, si los espacios n -dimensionales, e incluso los cuatridimensionales, parecen ficciones matemáticas sin sentido, lo mismo podría decirse de un espacio infinito-dimensional. Pero esto no es realmente así. Los razonamientos que se refieren a espacios de infinitas dimensiones no son difíciles en absoluto. Hoy en día

constituyen una importante rama de la matemática: el análisis funcional (ver capítulo 18).

Es de notar que la geometría analítica infinito-dimensional tiene aplicaciones muy importantes y es fundamental para la física contemporánea.

En cuanto a la geometría algebraica, es una continuación más directa de la geometría analítica ordinaria, que a su vez sólo es una parte de la geometría algebraica. Esta puede considerarse como la parte de la matemática que trata de las curvas, superficies e hipersuperficies, representadas en coordenadas cartesianas no sólo por ecuaciones algebraicas de primer y segundo grado, sino también de grados superiores. Se comprobará que en estas investigaciones es ventajoso considerar tanto coordenadas reales como coordenadas complejas, es decir, estudiar los problemas en un espacio llamado complejo. Los resultados más importantes en este campo fueron obtenidos en el siglo pasado por Riemann. Como brillante ejemplo de teoremas referentes a curvas de órdenes elevados, enunciamos un importante resultado de I. G. Petrovski sobre el número de óvalos en los que puede descomponerse una curva de orden n . Petrovski demostró que, si p es el número de óvalos que no están contenidos en otros óvalos o que están contenidos dentro de un número par de óvalos, y m es el número de óvalos contenidos dentro de un número impar de

óvalos, y considerando sólo curvas cuyas componentes ovals no se cortan a sí mismas ni a otras (Fig. 92), entonces

$$p - m \leq \frac{3n^2 - 6n}{8} + 1,$$

donde n es el orden de la curva, es decir, el grado de la ecuación que representa a la curva.

Este resultado es tanto más importante por cuanto hasta entonces no se sabía casi nada sobre la forma general de una curva de orden superior. Es sin duda uno de los teoremas más importantes de la geometría analítica.

Capítulo 4

ALGEBRA: TEORIA DE LAS ECUACIONES ALGEBRAICAS

§1. *Introducción*

Los rasgos característicos del álgebra son conocidos de todos, puesto que en el bachillerato se dan nociones elementales (aunque fundamentales) sobre ella. El álgebra se caracteriza, en primer lugar, por sus métodos, que implican el uso de letras y expresiones literales, sobre las cuales se realizan operaciones con propiedades dadas. En álgebra elemental las letras denotan números ordinarios, por lo que las leyes de las operaciones con expresiones literales se basan en las leyes generales de las operaciones con números. Por ejemplo, la suma no depende del orden de los sumandos, hecho que en álgebra se escribe así: $a + b = b + a$; para multiplicar la suma de dos números, podemos multiplicar cada uno de ellos por separado y sumar luego los productos obtenidos: $(a + b)c = ac + bc$, etc.

Si seguimos la demostración de un teorema algebraico, se ve que ésta sólo depende de las leyes u operaciones con números, y no de lo que representan las letras.

El método algebraico, o lo que es lo mismo, el método de cálculo literal, está presente en toda la matemática. De hecho, una parte

sustancial de la solución de un problema matemático resulta ser a menudo un simple cálculo algebraico más o menos complicado. Además, en la matemática se emplean distintos cálculos simbólicos en los que las letras ya no notan números sino otras entidades, y las propiedades de las operaciones sobre estas entidades pueden diferir de las propiedades del álgebra elemental. Por ejemplo, en geometría, mecánica y física se utilizan vectores y, como es bien sabido, las propiedades de las operaciones sobre los vectores son en parte las mismas que para los números y en parte esencialmente diferentes.

La importancia del método algebraico en la matemática moderna y su campo de aplicaciones han aumentado grandemente en las décadas recientes.

En primer lugar, las crecientes demandas de la tecnología obligan a reducir a resultados numéricos las soluciones de difíciles problemas del análisis matemático, y esto, generalmente, sólo es factible después de la algebrización de estos problemas, proceso que a su vez crea nuevos y complicados problemas dentro del álgebra misma.

En segundo lugar, ciertos problemas del análisis no resultaron claros y comprensibles hasta que fueron abordados por métodos algebraicos basados en una generalización profunda (al caso de un número infinito de incógnitas) de la teoría de los sistemas de ecuaciones de primer grado.

Finalmente, las partes más avanzadas del álgebra han encontrado aplicación en la física contemporánea: en efecto, los conceptos fundamentales de la mecánica cuántica se expresan en función de entidades algebraicas complicadas y nada elementales.

Los rasgos básicos de la historia del álgebra son los siguientes.

En primer lugar hay que hacer notar que las ideas sobre lo que el álgebra es y sobre cuál es su problema fundamental se han visto modificadas dos veces a lo largo de la historia: una en la primera mitad del siglo pasado, y otra al comienzo de nuestro siglo. Así, el álgebra ha significado en distintos momentos tres cosas bastante diferentes. En este aspecto la historia del álgebra difiere de las tres ramas más conocidas de la matemática: la geometría analítica, el cálculo diferencial y el cálculo integral; fueron elaboradas por sus creadores —Fermat, Descartes, Newton, Leibnitz y otros— y se desarrollaron y ampliaron rápidamente, a veces por adición de nuevas y extensas secciones, pero su carácter fundamental sufrió escasas modificaciones.

En la Antigüedad, cualquier ley que se descubría para resolver un determinado tipo de problemas matemáticos se expresaba senci-

llamente en palabras, puesto que los cálculos simbólicos aún no habían sido inventados. La palabra «álgebra» tuvo su origen en el título del importante trabajo del matemático kharizmiano del siglo IX Mohammed ibn Musa Al-Kharizmi (ver capítulo 1), en cuyos trabajos aparece la primera fórmula general para la resolución de las ecuaciones de primer y segundo grado. Sin embargo, la introducción de la notación simbólica se asocia generalmente con el nombre de Vieta, quien comenzó a notar por letras no sólo las incógnitas, sino también números dados. Descartes también contribuyó grandemente al desarrollo de la notación simbólica, y también, naturalmente, utilizó letras para representar números. En este momento el álgebra comienza a ser realmente la ciencia de los cálculos simbólicos, de las transformaciones de las fórmulas literales, de las ecuaciones algebraicas, etc., en contraste con la aritmética, que siempre opera sobre números concretos. Conceptos matemáticos intrincados se hicieron entonces claros y accesibles a la investigación, puesto que el concentrar la atención sobre la fórmula literal permite, en la mayor parte de los casos, ver su aspecto general o ley de formación y aplicarle transformaciones convenientes. En aquel tiempo, todo lo que en la matemática no era geometría ni análisis infinitesimal se llamaba álgebra. Esta fue la primera concepción, la de Vieta, en lo que toca al álgebra, concepción que quedó muy claramente expresada en el conocido libro «Introducción al Álgebra», escrito en el año 1760, es decir, hace doscientos años, por un miembro de la Academia Rusa de Ciencias, el famoso L. Euler.

Euler definió el álgebra como la teoría de los cálculos con cantidades de distintas clases. La primera parte de su libro contiene la teoría de los cálculos con los números racionales enteros, las fracciones ordinarias, las raíces cuadradas y cúbicas, la teoría de los logaritmos, las progresiones, la teoría del cálculo con polinomios y la teoría de la serie binómica de Newton y sus aplicaciones. La segunda parte consta de la teoría de las ecuaciones de primer grado y de los sistemas de tales ecuaciones, la teoría de las ecuaciones cuadráticas y de las soluciones de las ecuaciones de tercer y cuarto grado por radicales, así como una extensa sección sobre métodos de resolución de diversas ecuaciones indeterminadas con números enteros. Por ejemplo, se demostró que la ecuación de Fermat $x^3 + y^3 = z^3$ no tiene solución para x, y, z enteros.

A finales del siglo XVIII y comienzos del XIX, uno de los problemas del álgebra empezó a ocupar gradualmente el lugar central; era la teoría de la resolución de ecuaciones algebraicas, en la que la difi-

cultad fundamental estribaba en la solución de una ecuación algebraica de grado n con una incógnita:

$$x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_{n-1}x + a_n = 0.$$

Esto sucedió como consecuencia natural de la importancia de este problema para toda la matemática pura y aplicada, y también debido a la dificultad y profundidad de la mayoría de los teoremas relacionados con él.

La fórmula

$$x = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q},$$

para la solución de la ecuación de segundo grado $x^2 + px + q = 0$, era conocida de todos. Los algebristas italianos del siglo XVI encontraron reglas generales análogas, aunque más complicadas, para la solución de las ecuaciones de tercer y cuarto grado. Posteriores investigaciones en esta dirección en torno a las ecuaciones de grado superior tropezaron, sin embargo, con dificultades insalvables. Los más grandes matemáticos de los siglos XVI, XVII, XVIII y comienzos del XIX (Tartaglia, Cardano, Descartes, Newton, d'Alembert, Tschirnhausen, Bézout, Lagrange, Gauss, Abel, Galois, Lobachevski, Sturm y otros) crearon un impresionante edificio de teoremas y métodos relacionados con este problema. El álgebra en dos tomos de Serret (un trabajo que hizo época en su tiempo, puesto que presentó por primera vez lo más avanzado de la teoría de las ecuaciones algebraicas: la teoría de Galois) apareció a mediados del siglo XIX, exactamente cien años después del libro de Euler; en él ya se definía el álgebra como la teoría de las ecuaciones algebraicas. Este es el segundo punto de vista respecto al álgebra.

En la segunda mitad del siglo pasado tuvo lugar, sobre la base de las ideas de Galois en punto a la teoría de las ecuaciones algebraicas, un profundo desarrollo de la teoría de grupos¹ y de la teoría de los números algebraicos (cuya elaboración, en gran parte, se debió al matemático ruso E. I. Zolotarev).

También en este segundo período, y en relación con los citados problemas de resolución de una ecuación algebraica y con la teoría de las variedades algebraicas de orden superior (que por entonces se estudiaban en geometría analítica), se desarrolló el aparato alge-

¹ Ver capítulo 19.

braico en diferentes direcciones: por ejemplo, la teoría de los determinantes y matrices, la teoría algebraica de las formas cuadráticas y las transformaciones lineales y, en particular, la teoría de los invariantes. A lo largo de casi toda la segunda mitad del siglo XIX, la teoría de los invariantes constituyó un tema central del álgebra. A su vez, el desarrollo de la teoría de grupos y de la teoría de los invariantes tuvieron en este período una gran influencia sobre el desarrollo de la geometría ².

Un tercer punto de vista acerca de la naturaleza del álgebra surgió principalmente en conexión con lo siguiente. En la segunda mitad del siglo pasado, y dentro del campo de la mecánica, la física y la misma matemática, los científicos comenzaron a estudiar cada vez con mayor frecuencia objetos para los que era natural considerar operaciones de adición y substracción, y a veces multiplicación y división, pero para los cuales estas operaciones estaban sujetas a leyes muy distintas de las de los números racionales. Puesto que en una sección anterior hablamos ya de los vectores, nos limitaremos a mencionar ahora otras clases de objetos matemáticos con diferentes operaciones: matrices, tensores, espinores, números hipercomplejos. Todas estas magnitudes se notan por letras, pero sus operaciones difieren unas de otras. Si para un conjunto dado de objetos (notados por letras) se definen ciertas operaciones que satisfagan ciertas propiedades, se dice entonces que se ha definido un sistema algebraico.

El tercer punto de vista sobre el álgebra consiste en considerarla como el estudio de los diferentes sistemas algebraicos. Esta es la llamada álgebra axiomática o abstracta. Es abstracta porque en una etapa del cálculo no importa en absoluto lo que representen las letras en el sistema algebraico; lo único que interesa son los axiomas o leyes verificados por las operaciones; y se llama axiomática porque el álgebra está construida exclusivamente a partir de los axiomas establecidos al principio. Es como si se hubiese vuelto —aunque a nivel superior— al primer punto de vista (el de Vieta) sobre el álgebra, que considera a ésta como la teoría de los cálculos simbólicos. Aunque no tiene importancia lo que en sí noten las letras y sólo interesen las reglas operativas, no hay que olvidar, naturalmente, que de todos los sistemas algebraicos sólo son interesantes aquellos que tienen una gran significación, bien en la propia matemática, bien en sus aplicaciones.

² Ver capítulo 16.

La gran cantidad de material algebraico recogido en el período previo sirvió de base real para la construcción del álgebra abstracta contemporánea.

Los primeros años de la década de 1930-39 vieron la aparición del conocido libro de van der Waerden «Álgebra Moderna», que fue de gran importancia en la propagación de este punto de vista sobre el álgebra. El texto de A. G. Kurosch sobre el álgebra está orientado en la misma dirección.

En el siglo actual el álgebra ha encontrado profundas aplicaciones en la geometría (topología y teoría de los grupos de Lie) y, como ya dijimos, en la física contemporánea, especialmente en el análisis funcional y en la mecánica cuántica.

Particularmente importantes en la actualidad son los problemas de mecanización de los cálculos algebraicos por medio de distintas máquinas de cálculo, en especial las máquinas electrónicas de alta velocidad. Las cuestiones relacionadas con este tipo de matemática calculatoria están dando lugar a nuevos y diferentes problemas algebraicos.

En la presente obra hay dos capítulos (sin contar éste) dedicados al álgebra: el álgebra lineal (capítulo 15) y la teoría de grupos y otros sistemas algebraicos (capítulo 19).

§2. *Solución algebraica de una ecuación*

Una ecuación algebraica de grado n en una incógnita es una ecuación de la forma

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1} x + a_n = 0,$$

en la que $a_1, a_2, \dots, a_n$ son coeficientes conocidos³.

Ecuaciones de primer y segundo grado

Si la ecuación es de primer grado, tiene la forma

$$x + a = 0$$

y se resuelve inmediatamente

$$x = -a.$$

³ Se supone que todos los términos de la ecuación se pasan al primer miembro y que la ecuación se divide por el coeficiente de la potencia más alta de la incógnita.

La ecuación de segundo grado

$$x^2 + px + q = 0$$

fue ya resuelta en tiempos de la Antigüedad. Su solución es muy sencilla: si se pasa q , cambiando el signo, al segundo miembro, y se suma $p^2/4$ a ambos miembros de la ecuación, se tiene

$$x^2 + px + \frac{p^2}{4} = \frac{p^2}{4} - q.$$

Pero

$$x^2 + px + \frac{p^2}{4} = \left(x + \frac{p}{2}\right)^2;$$

Por tanto

$$x + \frac{p}{2} = \pm \sqrt{\frac{p^2}{4} - q},$$

de donde se obtiene la conocida fórmula para la solución de una ecuación de segundo grado:

$$x = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}.$$

La ecuación de tercer grado

Algo muy distinto sucedió con las ecuaciones de grado superior a dos. La ecuación general de tercer grado necesitó ya de consideraciones bastante profundas y resistió todos los esfuerzos de los matemáticos de la Antigüedad, logrando por fin resolverla, a principios del siglo XVI, durante el Renacimiento en Italia, el matemático italiano Scipio del Ferro. Del Ferro, siguiendo la costumbre de su tiempo, no publicó sus descubrimientos, sino que se los comunicó a uno de sus discípulos. Tras la muerte de Del Ferro, este alumno retó a uno de los grandes matemáticos italianos, llamado Tartaglia, proponiéndole resolviera un cierto número de ecuaciones de tercer grado. Tartaglia (1500-1557) aceptó el desafío, y ocho días antes de terminar el plazo encontró un método para resolver cualquier ecuación cúbica de la forma $x^3 + px + q = 0$.

En dos horas resolvió todos los problemas de su oponente. Un profesor de física y matemática de Milán, Cardano (1501-1576), sabedor de los descubrimientos de Tartaglia, comenzó a suplicarle que le

informase de su secreto. Tartaglia accedió finalmente, pero con la condición de que Cardano mantuviese su método en secreto. Cardano violó la promesa y publicó el resultado de Tartaglia en su trabajo «El arte sumo» (*Ars Magna*).

La fórmula para la resolución de una ecuación cúbica se ha llamado desde entonces fórmula de Cardano, aunque más correcto sería llamarla fórmula de Tartaglia.

La fórmula de Cardano se deduce como sigue.

En primer lugar, la solución de la ecuación cúbica general

$$y^3 + ay^2 + by + c = 0 \quad (1)$$

puede reducirse fácilmente a la solución de una ecuación cúbica de la forma

$$x^3 + px + q = 0, \quad (2)$$

que no contiene términos de segundo grado en la incógnita. Para lograr esto es suficiente poner $y = x - a/3$. En efecto, sustituyendo esta expresión en la ecuación (1) y quitando paréntesis, se obtiene

$$\left(x - \frac{a}{3}\right)^3 + a\left(x - \frac{a}{3}\right)^2 + b\left(x - \frac{a}{3}\right) + c = x^3 - 3x^2\frac{a}{3} + \cdots + ax^2 + \cdots,$$

donde los puntos suspensivos indican aquellos términos en los cuales x está elevada a la primera potencia o que son independientes de x . Los términos en x^2 se anulan entre sí.

Consideremos ahora la siguiente ecuación:

$$x^3 + px + q = 0.$$

Hacemos $x = u + v$, es decir, en lugar de una incógnita consideramos dos, u y v , transformando así el problema en otro con dos incógnitas. Se tiene

$$(u + v)^3 + p(u + v) + q = 0,$$

o bien

$$u^3 + v^3 + q + (3uv + p)(u + v) = 0.$$

Cualquiera que sea la suma de los dos números $u + v$, siempre es posible exigir que su producto uv tenga un valor fijado de antemano.

Si $u + v = A$, y exigimos que $uv = B$, entonces, puesto que $v = A - u$, se obtiene

$$u(A - u) = B,$$

por lo que es suficiente que u sea solución de la ecuación de segundo grado

$$u^2 - Au + B = 0,$$

y se sabe que toda ecuación de segundo grado tiene raíces reales o complejas, dadas por la fórmula antes deducida. En este caso, $u + v$ es igual a la raíz buscada x de la ecuación cúbica y se pide que

$$uv = -\frac{p}{3},$$

es decir, que $3uv + p = 0$. Con esta elección de u y v obtenemos

$$u^3 + v^3 + q = 0, \quad (3)$$

$$3uv + p = 0.$$

Por tanto, si encontramos los números u y v que satisfagan este sistema de ecuaciones, el número $x = u + v$ será raíz de la ecuación.

A partir del sistema (3) es fácil formar una ecuación de segundo grado cuyas raíces sean u^3 y v^3 . En efecto, se tiene

$$u^3 + v^3 = -q,$$

$$u^3 v^3 = -\frac{p^3}{27},$$

y, de acuerdo con un teorema utilizado anteriormente, u^3 y v^3 serán las raíces de la ecuación de segundo grado

$$z^2 + qz - \frac{p^3}{27} = 0.$$

Resolviéndola por la fórmula usual, se obtiene

$$u^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}, \quad v^3 = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}},$$

y, por tanto,

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}},$$

que es la fórmula de Cardano.

La ecuación de cuarto grado

Poco después de la resolución de la ecuación cúbica, Ferrari (1522-1565) resolvió la ecuación general de cuarto grado. Para la solución de la ecuación de tercer grado hemos visto que se necesita la solución preliminar de la ecuación auxiliar de segundo grado

$$z^2 + qz - \frac{p^3}{27} = 0,$$

donde $z = u^3$ o v^3 ; análogamente, la solución de una ecuación de cuarto grado puede basarse en la solución preliminar de una ecuación cúbica auxiliar.

El método de Ferrari consiste en lo siguiente. Sea la ecuación general de cuarto grado

$$x^4 + ax^3 + bx^2 + cx + d = 0.$$

Escribámosla de nuevo en la forma:

$$x^4 + ax^3 = -bx^2 - cx - d$$

y sumemos a ambos miembros $a^2x^2/4$; el primer miembro es entonces un cuadrado perfecto:

$$\left(x^2 + \frac{ax}{2}\right)^2 = \left(\frac{a^2}{4} - b\right)x^2 - cx - d.$$

Sumando a ambos miembros de la ecuación los términos

$$\left(x^2 + \frac{ax}{2}\right)y + \frac{y^2}{4}$$

(donde y es una nueva variable, a la que se impondrá más tarde una condición necesaria), se obtiene en el primer miembro un cuadrado perfecto:

$$\left(x^2 + \frac{ax}{2} + \frac{y}{2}\right)^2 = \left(\frac{a^2}{4} - b + y\right)x^2 + \left(\frac{ay}{2} - c\right)x + \left(\frac{y^2}{4} - d\right). \quad (4)$$

El problema ha quedado reducido a otro con dos incógnitas.

En el segundo miembro de la ecuación (4) se tiene un trinomio de segundo grado en x , cuyos coeficientes dependen de y . Elijamos y de modo que este trinomio sea el cuadrado de un binomio de primer grado $\alpha x + \beta$.

Para que el trinomio de segundo grado $Ax^2 + Bx + C$ sea el cuadrado del binomio $\alpha x + \beta$, es suficiente que

$$B^2 - 4AC = 0.$$

En efecto, si $B^2 - 4AC = 0$, entonces

$$Ax^2 + Bx + C = (\sqrt{A}x + \sqrt{C})^2,$$

es decir,

$$Ax^2 + Bx + C = (\alpha x + \beta)^2,$$

donde

$$\alpha = \sqrt{A}, \quad \beta = \sqrt{C}.$$

Por tanto, si se elige y tal que

$$\left(\frac{ay}{2} - c\right)^2 - 4\left(\frac{a^2}{4} - b + y\right)\left(\frac{y^2}{4} - d\right) = 0,$$

la primera parte de la ecuación (4) será el cuadrado perfecto $(\alpha x + \beta)^2$. Quitando paréntesis, se obtiene una ecuación en y :

$$y^3 - by^2 + (ac - 4d)y - [d(a^2 - 4b) + c^2] = 0.$$

Resolviendo esta ecuación cúbica auxiliar (por ejemplo, por la fórmula de Cardano) se calculan α y β en función de su solución y_0 , es decir,

$$\left(x^2 + \frac{ax}{2} + \frac{y_0}{2}\right)^2 = (\alpha x + \beta)^2,$$

de donde

$$x^2 + \frac{ax}{2} + \frac{y_0}{2} = ax + \beta, \quad \text{o bien} \quad x^2 + \frac{ax}{2} + \frac{y_0}{2} = -ax - \beta.$$

A partir de estas dos ecuaciones de segundo grado se pueden encontrar las cuatro raíces de la ecuación de cuarto grado dada.

Así es como los matemáticos italianos del siglo XVI resolvieron las ecuaciones algebraicas de tercer y cuarto grado.

El éxito de los matemáticos italianos produjo un gran efecto, pues fue la primera ocasión en que la ciencia moderna superaba los logros de la antigua. Hasta entonces, en el transcurso de la Edad Media, el único propósito había sido el de entender el trabajo de los antiguos; ahora, por el contrario, quedaban resueltas ciertas cuestiones que los antiguos no habían podido tratar con éxito. Y esto sucedía en los años 1500-09, es decir, un siglo antes de la invención de las nuevas ramas de la matemática: la geometría analítica, el cálculo diferencial y el cálculo integral, que dejaron sentada finalmente la superioridad de la nueva ciencia sobre la antigua. Después de esto no hubo matemático importante que no intentara extender los logros de los italianos y resolver las ecuaciones de quinto, sexto, y grados superiores de un modo análogo por medio de radicales.

El eminente algebrista del siglo XVII Tschirnhausen (1651-1708) creyó incluso que había encontrado finalmente un método general de resolución. Su método se basaba en la transformación de una ecuación en otra más sencilla, pero que necesitaba de la solución de ecuaciones auxiliares. Posteriormente, mediante un análisis más profundo, se demostró que el método de Tschirnhausen daba en efecto la solución de las ecuaciones de tercer y cuarto grado, pero para una ecuación de quinto grado ya necesitaba la resolución previa de una ecuación auxiliar de sexto grado, cuyas soluciones eran a su vez desconocidas.

Factorización de un polinomio y fórmulas de Vieta

Si se acepta sin demostración el llamado teorema fundamental del álgebra ⁴, que dice que toda ecuación

$$f(x) = 0,$$

⁴ La demostración del teorema fundamental del álgebra es difícil y se dio considerablemente más tarde. Dedicaremos el §3 a ella. Pero su validez se supuso mucho antes de que fuese rigurosamente demostrado.

donde

$$f(x) = x^n + a_1x^{n-1} + \cdots + a_n$$

es un polinomio en x de grado n y los coeficientes $a_1, a_2, \dots, a_n$ son números reales o complejos dados, tiene al menos una raíz real o compleja, y teniendo en cuenta que todos los cálculos con números complejos siguen las mismas reglas que con números reales, entonces es fácil demostrar que el polinomio $f(x)$ puede representarse (y de manera única) como producto de factores de primer grado

$$f(x) = (x - a)(x - b) \cdots (x - l),$$

donde $a, b, \dots, l$ son números reales o complejos.

En efecto, sea n una raíz de $f(x)$; dividamos $f(x)$ por $x - a$; puesto que el divisor es de primer grado, el resto será un número constante R , es decir, tendremos la identidad

$$f(x) = (x - a)f_1(x) + R$$

donde $f_1(x)$ es un polinomio de grado $n - 1$ y R una constante. Sustituyendo en lugar de x el número a , obtenemos

$$f(a) = (a - a)f_1(a) + R = R.$$

Pero, puesto que a es raíz de $f(x)$, tenemos $f(a) = 0$, y por tanto $R = 0$; es decir, un polinomio puede siempre dividirse por $(x - a)$ dando resto cero, si a es una raíz de este polinomio. Así

$$f(x) = (x - a)f_1(x).$$

Pero, si admitimos la validez del teorema fundamental del álgebra, el polinomio $f_1(x)$ tendrá una raíz b , y obtenemos análogamente

$$f_1(x) = (x - b)f_2(x),$$

donde el polinomio es ya de grado $(n - 2)$, etc. Esta factorización, como puede demostrarse fácilmente, es única.

Todo polinomio $f(x)$ de grado n tiene, pues, n y sólo n raíces $a, b, \dots, l$. Estas raíces pueden ser todas distintas o puede suceder que algunas de ellas sean iguales. En este segundo caso se dice que

la raíz correspondiente del polinomio $f(x)$ es una raíz múltiple (doble, triple, etc.).

Desarrollando la expresión

$$(x - a)(x - b)(x - c) \cdots (x - l)$$

y comparando los coeficientes de iguales potencias de x , vemos inmediatamente que

$$-a_1 = a + b + c + \cdots + l,$$

$$a_2 = ab + ac + \cdots + kl,$$

$$-a_3 = abc + abd + \cdots,$$

$$\cdots \cdots \cdots l,$$

$$\pm a_n = abc \cdots l,$$

que son las fórmulas de Vieta.

Un teorema sobre polinomios simétricos

Las fórmulas de Vieta son polinomios en las n letras $a, b, \dots, l$ que no varían al permutar dichas letras. En efecto, $a + b + \cdots + k + l = b + a + \cdots + k + l$, etc. En general, tales polinomios en n letras, que no varían por permutación de éstas, se llaman polinomios simétricos en n letras. Por ejemplo, $5x^2 + 5y^2 - 7xy$ es un polinomio simétrico en x e y . Es posible demostrar el teorema de que todo polinomio simétrico entero en n letras con coeficientes arbitrarios $A, B, \dots$ puede expresarse en forma racional entera (es decir, mediante operaciones de adición, sustracción y multiplicación) en función de los coeficientes $A, B, \dots$ y de los polinomios de Vieta. Si $a, b, \dots, l$ son las raíces de una ecuación de grado n , $x^n + a_1x^{n-1} + \cdots + a_n = 0$, entonces todo polinomio simétrico en $a, b, \dots, l$ con coeficientes arbitrarios $A, B, \dots$ puede ser expresado en forma racional entera en función de estos coeficientes $A, B, \dots$ y de los coeficientes $a_1, a_2, \dots, a_n$ de la ecuación. Este es el llamado teorema fundamental de los polinomios simétricos.

Contribuciones de Lagrange

En su gran trabajo «Reflexiones sobre la resolución de ecuaciones algebraicas», publicado en 1770-1771 (con más de doscientas páginas), el famoso matemático francés Lagrange examinó críticamente todas las soluciones de las ecuaciones de segundo, tercer y cuarto grado

que se conocían hasta entonces, y demostró que dichas soluciones se basaban siempre en propiedades que no se verificaban para las ecuaciones de grado quinto o superior.

Desde los tiempos de Del Ferro hasta este trabajo de Lagrange habían pasado más de dos siglos y medio sin que nadie, durante este largo período, hubiese dudado de la posibilidad de resolver las ecuaciones de quinto grado y superiores por el método de radicales, es decir, de encontrar ciertas fórmulas que implicaran sólo las operaciones de adición, sustracción, multiplicación, división y radicación con exponentes enteros positivos, y mediante las cuales se pudiese expresar la solución de una ecuación en función de sus coeficientes: en resumen, fórmulas similares a las que se habían encontrado en tiempos de la Antigüedad para la ecuación de segundo grado, y para las de tercer y cuarto grado, a principios del siglo XVI, por los italianos. Esta situación se atribuía generalmente a la incapacidad de los matemáticos para encontrar una solución que, aunque válida, se hallaba muy escondida.

Lagrange dice en su memoria: «El problema de resolver (por radicales) ecuaciones de grado mayor que cuatro es uno de aquellos problemas que no han sido resueltos, aunque nada demuestra la imposibilidad de su resolución»; y dos páginas más adelante añade esto otro: «Por nuestro razonamiento vemos que es muy dudoso que los métodos que hemos considerado puedan dar una solución completa de las ecuaciones de quinto grado».

En sus investigaciones Lagrange introdujo la expresión

$$a + \epsilon b + \epsilon^2 c + \cdots + \epsilon^{n-1} l$$

en las raíces $a, b, \dots, l$ de una ecuación, donde ϵ es una raíz n -ésima de la unidad ⁵, habiendo establecido que tales expresiones están inti-

⁵ Es decir, un número complejo que elevado a la potencia n -ésima es igual a uno. Por ejemplo, la raíces cúbicas de la unidad son:

$$1, -\frac{1}{2} + \frac{\sqrt{3}}{2}i, -\frac{1}{2} - \frac{\sqrt{3}}{2}i,$$

donde $i = \sqrt{-1}$ (ver §3). En efecto:

$$\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)^3 = -\frac{1}{8} - \frac{3}{8}\sqrt{3}i + \frac{9}{8} + \frac{3\sqrt{3}}{8}i = 1,$$

y análogamente

$$\left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)^3 = 1.$$

mamente relacionadas con la solución de las ecuaciones por radicales. Estas expresiones se llaman ahora «resolventes de Lagrange».

Además, Lagrange observó que la teoría de las permutaciones de las raíces de una ecuación es de gran importancia en la resolución de ecuaciones. Incluso expresó la idea de que la teoría de las permutaciones es la «verdadera filosofía de toda la cuestión», en lo cual acertó plenamente, como demostraron las investigaciones posteriores de Galois.

El método de Lagrange para la resolución de las ecuaciones de segundo, tercero y cuarto grados no era el mismo que el de los italianos, que siempre se basaba en transformaciones especiales de una clase complicada y, por así decirlo, accidental. Los métodos de Lagrange estaban completamente ordenados y desarrollados a partir de una idea general en la que intervenía la teoría de los polinomios simétricos, la teoría de las permutaciones y la teoría de resolventes.

Consideremos, por ejemplo, la solución por el método de Lagrange de la ecuación general de cuarto grado

$$x^4 + mx^3 + nx^2 + px + q = 0.$$

Sean a, b, c, d las raíces de esta ecuación. Consideremos la resolvente

$$a + b - c - d,$$

es decir,

$$a + \epsilon c + \epsilon^2 b + \epsilon^3 d,$$

donde $\epsilon = -1$. Si permutamos a, b, c, d de todas las formas posibles ($1 \cdot 2 \cdot 3 \cdot 4 = 24$ formas en total) obtenemos en total seis expresiones diferentes;

$$\begin{aligned} a + b - c - d, \\ a + c - b - d, \\ a + d - c - b, \\ c + d - a - b, \\ b + d - a - c, \\ b + c - a - d. \end{aligned} \tag{5}$$

Una ecuación de sexto grado cuyas raíces sean estas seis expresiones tendrá así coeficientes que no varían en las 24 permutaciones de a, b, c, d , puesto que cualquiera de las 24 permutaciones puede

sólo permutar estas expresiones entre ellas mismas y los coeficientes de la ecuación de sexto grado no dependen del orden en el que tomemos sus raíces. Así, estos coeficientes son polinomios simétricos en a, b, c, d . Pero, en virtud del teorema fundamental de los polinomios simétricos, estos coeficientes se expresan en forma racional entera en función de los coeficientes m, n, p, q de la ecuación. Además, puesto que las expresiones (5) son opuestas dos a dos, esta ecuación de sexto grado contendrá únicamente términos en potencias pares. En efecto, notando las expresiones (5) por $\alpha, \beta, \gamma, -\alpha, -\beta, -\gamma$, respectivamente, el primer miembro de la ecuación de sexto grado será igual a

$$(y - \alpha)(y + \alpha)(y - \beta)(y + \beta)(y - \gamma)(y + \gamma) = \\ = (y^2 - \alpha^2)(y^2 - \beta^2)(y^2 - \gamma^2),$$

de donde obtenemos la ecuación de sexto grado

$$y^6 - (3m^2 - 8n)y^4 + 3(m^4 - 16m^2n - 16n^2 + 16mp - 64q)y^2 - \\ - (m^2 - 4m + 8p)^2 = 0.$$

Poniendo $y^2 = t$, obtenemos una ecuación cúbica en t ; y si t', t'', t''' son raíces, entonces

$$a + b - c - d = \sqrt{t'}, \\ a + c - b - d = \sqrt{t''}, \\ a + d - b - c = \sqrt{t'''},$$

Tenemos también

$$a + b + c + d = -m.$$

Sumando estas ecuaciones después de multiplicarlas por $1, 1, 1, 1$; $1, -1, -1, 1$; $-1, 1, -1, 1$; y $-1, -1, 1, 1$, obtenemos

$$a = \frac{1}{4} (-m + \sqrt{t'} + \sqrt{t''} + \sqrt{t'''}), \\ b = \frac{1}{4} (-m + \sqrt{t'} - \sqrt{t''} - \sqrt{t'''}), \\ c = \frac{1}{4} (-m - \sqrt{t'} + \sqrt{t''} - \sqrt{t'''}), \\ d = \frac{1}{4} (-m - \sqrt{t'} - \sqrt{t''} + \sqrt{t'''}).$$

Así, la solución de una ecuación de cuarto grado se reduce a la solución de una ecuación cúbica; y las de segundo y tercer grado se resuelven análogamente.

Lagrange dio gran impulso a la teoría de las ecuaciones algebraicas. Sin embargo, y a pesar de sus persistentes esfuerzos, el problema de la solución por radicales de las ecuaciones algebraicas de grado superior a 4 permaneció como antes. Este problema, sobre el cual habían trabajado los matemáticos en vano durante casi tres siglos, constituía, en expresión de Lagrange, «un reto a la mente humana».

Descubrimiento de Abel

Fue, por consiguiente, una gran sorpresa para todos los matemáticos la publicación en el año 1824 del trabajo del joven genio noruego Abel (1802-1829); en este trabajo se daba la demostración de que, si los coeficientes $a_1, a_2, \dots, a_n$ de una ecuación se consideran únicamente como letras, no existe entonces ninguna expresión radical en estos coeficientes que sea raíz de la ecuación correspondiente, si su grado es $n \geq 5$. Los esfuerzos que durante tres siglos hicieron los más grandes matemáticos de todos los países para resolver las ecuaciones de grado quinto o superior por radicales no tuvieron éxito por la sencilla razón de que este problema no tiene solución.

Tal fórmula se conoce para las ecuaciones de segundo grado, y como vimos existen fórmulas análogas para las ecuaciones de tercer y cuarto grado, pero para las ecuaciones de grado 5 o mayor no existen.

La demostración de Abel es difícil y no se dará aquí.

Teoría de Galois

Pero esto no fue todo. Quedaba todavía llegar a un resultado muy importante en la teoría de las ecuaciones algebraicas. La cuestión es que hay muchas formas especiales de ecuaciones de cualquier grado que son resolubles por radicales, y muchas de ellas son precisamente ecuaciones de gran importancia en las aplicaciones. Tales son, por ejemplo, las del tipo $x^n = A$. Abel encontró otra clase muy amplia de tales ecuaciones, las llamadas ecuaciones cíclicas y las ecuaciones «abelianas». En conexión con el problema de la construcción con regla y compás de polígonos regulares, Gauss consideró explícitamente las ecuaciones ciclotómicas, es decir, ecuaciones de la forma

$$x^{p-1} + x^{p-2} + \dots + x + 1 = 0,$$

donde p es un número primo, y demostró que podían ser siempre reducidas a una cadena de ecuaciones de grado inferior; además, encontró condiciones necesarias y suficientes para que una tal ecuación pudiera ser resuelta por raíces cuadradas. La necesidad de estas condiciones sólo fue demostrada rigurosamente por Galois.

Así, después de Abel la situación era la siguiente: aunque —como este matemático demostró— la ecuación general de grado superior a 4 no puede resolverse por radicales, hay muchas ecuaciones especiales de grado arbitrario que pueden resolverse por radicales. La cuestión de la resolución de ecuaciones por radicales quedó colocada, a raíz de estos descubrimientos, sobre una base completamente nueva. Se vio claro que la tarea consistía en determinar exactamente qué ecuaciones podrían resolverse por radicales o, en otras palabras, cuáles eran las condiciones necesarias y suficientes para la resolución de una ecuación por radicales. Este problema, cuya solución aclaró, en cierta medida, todo el problema, fue resuelto por el ingenioso matemático francés Evaristo Galois.

Galois (1811-1832) pereció a la edad de veinte años en un duelo. En los dos últimos años de su vida no pudo dedicar mucho tiempo a la matemática, pues se vio arrastrado por el torbellino de la vida política durante la revolución de 1830, languideciendo en la cárcel por su discurso contra el régimen reaccionario de Luis Felipe. Sin embargo, y a pesar de su corta vida, Galois hizo descubrimientos muy adelantados respecto a los de su época en distintas partes de la matemática, y en particular obtuvo resultados muy importantes en la teoría de las ecuaciones algebraicas. En una pequeña publicación, «Memoria sobre las condiciones de resolubilidad de ecuaciones por radicales» —que permaneció en forma manuscrita después de su muerte y fue publicada por primera vez por Liouville en 1846—, Galois, partiendo de algunos conceptos muy sencillos aunque profundos, desentrañó finalmente el complejo conjunto de dificultades con el que los más sobresalientes matemáticos se debatieron sin éxito hasta su época. El éxito de Galois se basó en el hecho de que por primera vez introdujo en la teoría de ecuaciones una serie de nuevos conceptos generales sobremanera importantes, que fueron desde entonces de gran importancia en toda la matemática.

Consideremos la teoría de Galois en un caso especial, aquel en el que los coeficientes, $a_1, a_2, \dots, a_n$ de una ecuación dada de grado n

$$x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \quad (6)$$

son números racionales. Este caso es particularmente interesante y contiene ya en esencia todas las dificultades de la teoría general de Galois. Supondremos también que las raíces de esta ecuación $a, b, c, \dots$ son distintas.

Galois comenzó, al igual que Lagrange, considerando una expresión de primer grado en $a, b, c, \dots$

$$V = Aa + Bb + Cc + \dots,$$

aunque no impuso la condición de que los coeficientes $A, B, C, \dots$ de esta expresión fuesen raíces de la unidad, sino que tomó para $A, B, C, \dots$ números racionales enteros cualesquiera que dieran valores numéricamente distintos para todas las $n! = 1 \cdot 2 \cdot 3 \cdots n$ cantidades $V, V', V'', \dots, V^{(n!-1)}$ obtenidas a partir de V permutando las raíces $a, b, c, \dots$ de todos los $n!$ modos posibles. Esto siempre puede hacerse. Galois construye entonces la ecuación de grado $n!$, cuyas raíces son $V, V', V'', \dots, V^{(n!-1)}$. El teorema sobre polinomios simétricos demuestra que los coeficientes de esta ecuación $\phi(x) = 0$ de grado $n!$ serán números racionales.

Hasta ahora todo es muy parecido a lo que hizo Lagrange.

A continuación Galois introduce el primer concepto nuevo importante: el de irreducibilidad de un polinomio dentro de un cuerpo dado de números. Si tenemos un polinomio en x cuyos coeficientes, por ejemplo, son números racionales, el polinomio se dirá reducible en el cuerpo de los números racionales si se puede representar en forma de producto de polinomios de grados inferiores con coeficientes racionales. En caso contrario, el polinomio se dice irreducible en el cuerpo de los racionales. El polinomio $x^3 - x^2 - 4x - 6$ es reducible en el cuerpo de los racionales, puesto que es igual a $(x^2 + 2x + 2)(x - 3)$; en cambio, el polinomio $x^3 + 3x^2 + 3x - 5$ es irreducible en este mismo cuerpo.

Existen métodos (aunque requieren muchos cálculos) para factorizar cualquier polinomio de coeficientes racionales en polinomios irreducibles en el cuerpo de los racionales.

Galois factoriza luego el polinomio $\phi(x)$ en factores irreducibles en el cuerpo de los racionales.

Sea $F(x)$ uno de estos polinomios irreducibles (su elección particular no afecta a lo que sigue), y supongámoslo de grado m .

El polinomio $F(x)$ será entonces el producto de m de los factores de primer grado $x - V, x - V', \dots, x - V^{(m!-1)}$ en que se descompuso el polinomio $\phi(x)$ de grado $n!$. Sean $x - V, x - V', \dots,$

$x - V^{(m-1)}$ estos m factores. Enumeremos en cualquier orden las raíces $a, b, c, \dots, l$ de la ecuación de grado n dada (6), asignándoles los índices, $1, 2, \dots, n$. Entonces las $V, V', \dots, V^{(n!-1)}$ corresponden a las $n!$ posibles permutaciones de los números $1, 2, \dots, n$, correspondientes a las permutaciones de las raíces, y las $V, V', \dots, V^{(m-1)}$ corresponden a sólo m permutaciones. El conjunto G de estas m permutaciones de los números $1, 2, \dots, n$ se llama grupo de Galois de la ecuación (6) dada*.

A continuación Galois introduce nuevos conceptos y desarrolla razonamientos sencillos, aunque realmente notables, con los cuales demuestra que una condición necesaria y suficiente para la resolución de la ecuación (6) por radicales es que el grupo G de las permutaciones de los números $1, 2, \dots, n$ satisfaga cierta condición.

Fue así como la profecía de Lagrange, de que todo el problema reposaba sobre la teoría de las permutaciones, demostró ser cierta.

En particular, la demostración del teorema de Abel sobre la no resolubilidad de una ecuación general de quinto grado por radicales puede hacerse ahora como sigue. Puede demostrarse que existe un número arbitrario de ecuaciones de quinto grado, incluso de coeficientes racionales enteros, para las cuales el polinomio correspondiente $\phi(x)$ de grado 120 es irreducible, es decir, cuyo grupo de Galois es el grupo de las $5! = 120$ permutaciones de los índices $1, 2, 3, 4, 5$ de sus raíces. Pero este grupo, como puede demostrarse, no satisface el criterio de Galois, y por tanto estas ecuaciones de quinto grado no pueden resolverse por radicales. Por ejemplo, puede demostrarse que la ecuación $x^5 + x - a = 0$, donde a es un número entero positivo, no puede resolverse, en la mayoría de los casos, por radicales. No es resoluble por radicales para $a = 3, 4, 5, 7, 8, 9, 10, 11, \dots$.

La aplicación de la teoría de Galois al problema de la resolubilidad de los problemas geométricos con regla y compás

Una de las aplicaciones más interesantes de la teoría de Galois es la siguiente. Muchos problemas de geometría plana pueden resolverse por construcciones en las que se utiliza únicamente la regla y el compás. Por ejemplo, con regla y compás pueden construirse triángulos, cuadrados, pentágonos, exágonos, octógonos, decágonos, etc., regulares, pero es imposible construir un polígono regular de 7, 9 u 11 lados. ¿Qué problemas pueden resolverse con regla y compás y cuáles no? Antes de Galois era un problema no resuelto. La teoría de Galois proporciona la siguiente respuesta.

* Sobre grupos de Galois diremos más en §5, capítulo 19.

La solución simultánea de ecuaciones de dos rectas, una recta y una circunferencia o dos circunferencias puede reducirse a la solución de ecuaciones de primer o segundo grado. Esto es inmediato para el caso de una recta y una circunferencia; y en el caso de dos circunferencias $(x - a_1)^2 + (y - b_1)^2 = r_1^2$ y $(x - a_2)^2 + (y - b_2)^2 = r_2^2$, restando una ecuación de la otra desaparecen los términos x^2 e y^2 y se obtiene una ecuación de primer grado, a resolver simultáneamente con la de una de las circunferencias, con lo cual se tiene otra vez una ecuación de segundo grado. Por tanto, cada una de las etapas de un problema a resolver con regla y compás se reduce a una ecuación de primer o segundo grado y, por consiguiente, todos los problemas resolubles con regla y compás se reducen a una ecuación algebraica con una incógnita, cuya solución implica la extracción de una cadena de raíces cuadradas. Recíprocamente, si la solución de un problema geométrico se reduce a la de una ecuación algebraica de ese tipo, puede ser resuelto con regla y compás, puesto que las raíces cuadradas, como se sabe, pueden construirse con regla y compás.

Dado un problema geométrico, hay que plantear en primer lugar una ecuación algebraica equivalente al problema dado. Si es imposible plantear tal ecuación, el problema, evidentemente, no es resoluble con regla y compás. En el caso de que se logre plantear la ecuación, hay que seleccionar el factor irreducible que se halle relacionado con la solución del problema y determinar si esta ecuación irreducible se puede resolver por extracción de raíces cuadradas. Como demuestra la teoría de Galois, para ello es necesario y suficiente que el número m de permutaciones que constituye su grupo de Galois sea una potencia de 2.

Con este criterio podemos demostrar el teorema establecido por Gauss de que un polígono regular con un número primo p de lados se puede construir con regla y compás si, y sólo si, el número primo p es de la forma $2^{2^n} + 1$, es decir, para $p = 3, 5, 17, 257$, pero no para $p = 7, 11, 13, 19, 23, 29, 31, \dots$, etc. Gauss demostró sólo la parte necesaria de este aserto, pero no la suficiente.

Por el mismo método se puede demostrar que es imposible dividir un ángulo arbitrario en tres partes iguales con regla y compás; ni duplicar el cubo, es decir, dada la arista de un cubo encontrar la arista de un cubo de volumen doble, etc.

La imposibilidad de cuadrar el círculo —de construir con regla y compás el lado de un cuadrado de igual área que un círculo de radio dado— se demuestra de un modo diferente. Puede demostrarse que el lado de dicho cuadrado no está relacionado con el radio por una

ecuación algebraica, sino que es, por así decirlo, transcendente respecto al radio, y por tanto no es *a fortiori* expresable en función del radio a través de una cadena de raíces cuadradas. Esta demostración es difícil y no se deduce de la teoría de Galois.

Dos problemas fundamentales no resueltos relacionados con la teoría de Galois

En la teoría de Galois existen todavía dos problemas básicos que no han sido resueltos en forma general, aunque han sido muchos los matemáticos que han trabajado sobre ellos casi ininterrumpidamente.

El primero es el problema de las llamadas resolventes de Hilbert-Chebotarev (que no debe ser confundido con el de las resolventes de Lagrange), generalización directa del problema de la solución de ecuaciones por radicales. La idea es ésta: decir que una ecuación es resoluble por radicales es exactamente lo mismo que decir que su solución se reduce a una cadena de sucesivas ecuaciones binomias,

puesto que el radical $\sqrt[n]{A}$ es una raíz de la ecuación binomia $x^n = A$. Pero puede suceder que, aunque la ecuación no se pueda reducir a una cadena de ecuaciones tan sencillas como las binomias, sí sea posible reducirla a una cadena de ecuaciones muy sencillas de otro tipo. Antes de finalizar el siglo XVIII se había demostrado que la ecuación general de quinto grado puede ser reducida a una cadena de ecuaciones binomias junto con una ecuación de la forma $x^5 + x + A = 0$, que no es binomia, pero que tiene, al igual que las binomias, un único parámetro A .

Más tarde se demostró que las ecuaciones de sexto grado no pueden ser reducidas a una cadena de ecuaciones con un sólo parámetro. Para ecuaciones de grado arbitrario se trata de resolver el siguiente problema: ¿qué clase de ecuaciones más sencillas, es decir, con un número mínimo de parámetros, componen la cadena a la cual se puede reducir la ecuación?

Si la ecuación dada se reduce a una cadena de ecuaciones de un sólo parámetro, entonces para cada una de estas ecuaciones uniparamétricas se puede calcular una tabla que exprese sus raíces en función del parámetro. La solución de la ecuación dada se reduce así al uso de una cadena de tablas de esta índole.

En segundo lugar, un problema todavía más profundo consiste en el recíproco de la teoría de Galois. Galois demostró que las propiedades de las soluciones de una ecuación dependen de su grupo. Pero, recíprocamente, ¿puede cualquier grupo de permutaciones ser

grupo de Galois de alguna ecuación y pueden plantearse todas las ecuaciones cuyo grupo de Galois es uno dado?

Para la primera de estas dos cuestiones sólo se conocen resultados parciales, aunque matemáticos tan relevantes como Klein e Hilbert trabajaron sobre ello persistentemente; los primeros teoremas generales los dio el importante algebrista soviético H. G. Chebotarev.

La segunda cuestión, referente a los llamados grupos resolubles, es decir, grupos que satisfacen el criterio de Galois, fue resuelta en sentido afirmativo en años recientes por el matemático soviético I. R. Safarevich.

§3. *El teorema fundamental del álgebra*

En la sección anterior mencionamos ya los intentos que durante tres siglos se llevaron a cabo para resolver por radicales una ecuación de n -ésimo grado. El problema resultó ser muy difícil y profundo y condujo a la aparición de nuevos conceptos, importantes no sólo para el álgebra, sino para toda la matemática. En lo referente a la resolución práctica de ecuaciones, el resultado de todo este trabajo fue el siguiente: quedó claro que no era posible la solución por radicales de la mayoría de las ecuaciones algebraicas, e incluso siéndolo, era de poco valor práctico debido a su complejidad, excepto en el caso de la ecuación de segundo grado.

En vista de ello, hace ya tiempo que los matemáticos comenzaron a trabajar sobre la teoría de las ecuaciones algebraicas en tres direcciones diferentes, a saber: 1) sobre el problema de la existencia de una raíz; 2) sobre el problema de cómo deducir, a partir de los coeficientes de una ecuación, ciertas propiedades de sus raíces sin resolverla (por ejemplo, si tiene raíces reales y cuántas); y finalmente, 3) sobre el cálculo aproximado de las raíces de una ecuación.

En primer lugar, fue necesario demostrar que toda ecuación algebraica de grado n con coeficientes reales o complejos tiene al menos una raíz real o compleja ⁷.

Este teorema, que es uno de los más importantes de toda la matemática, careció durante mucho tiempo de una demostración rigurosa. En vista de su importancia y dificultad, se le llama generalmente «teorema fundamental del álgebra», aunque la mayoría de los

⁷ Lo importante es que existen ecuaciones algebraicas, por ejemplo, $a^x = 0$, que no tienen raíces, ni reales ni complejas.

métodos por los que ha sido demostrado están tan íntimamente relacionados con el análisis infinitesimal como con el álgebra. La primera demostración la dio d'Alembert, pero más tarde se comprobó que aquélla no era del todo completa. D'Alembert aceptó como trivial la proposición del análisis de que una función continua, definida sobre un conjunto de puntos cerrado y acotado, alcanza en alguno de los puntos un mínimo. Esto es cierto, pero tenía que ser demostrado. La demostración rigurosa de esta propiedad no llegó hasta la segunda mitad del siglo XIX, es decir, cien años después de las investigaciones de d'Alembert.

En general se admite que la primera demostración rigurosa del teorema fundamental del álgebra la dio Gauss; sin embargo, alguna de sus demostraciones parciales requieren, para un completo rigor, no menos añadidos que los que requería la demostración de d'Alembert. Hoy se conocen varias demostraciones diferentes, y completamente rigurosas, de este teorema.

En esta sección desarrollaremos la demostración del teorema fundamental del álgebra basada en el llamado lema de d'Alembert, y daremos también una demostración completa de la antedicha proposición del análisis.

La teoría de los números complejos

Antes de considerar la demostración del teorema fundamental del álgebra, debemos recordar en primer lugar la teoría de los números complejos tal como se estudia en el bachillerato. Las dificultades que condujeron a la creación de la teoría de los números complejos surgen por primera vez al tratar de resolver las ecuaciones de segundo grado. ¿Qué hacer cuando el número $p^2/4 - q$, bajo el signo radical en la fórmula para la solución de la ecuación de segundo grado, es negativo? No existe ningún número real, positivo o negativo, que sea raíz cuadrada de un número negativo, puesto que el cuadrado de cualquier número real es o positivo o cero.

Después de muchas dudas, que duraron más de un siglo, los matemáticos llegaron a la conclusión de que era necesario introducir un nuevo tipo de números, los llamados números complejos, con las siguientes operaciones.

Convencionalmente se introduce un número de nuevo carácter: $i = \sqrt{-1}$, tal que $i^2 = -1$, y se consideran números de la forma $a + bi$, donde a y b son números reales ordinarios. Los números $a + bi$ se llaman complejos. Dos de tales números, $a + bi$ y $c + di$,

se suponen iguales si $a = c$ y $b = d$. La suma de estos dos números se define como el número $(a + c) + (b + d)i$, y su diferencia es el número $(a - c) + (b - d)i$. Para multiplicarlos se procede como si fueran binomios, aunque teniendo en cuenta que $i^2 = -1$, es decir,

$$(a + bi)(c + di) = ac + bci + adi + bdi^2 = (ac - bd) + (bc + ad)i.$$

Si a y b se consideran como coordenadas rectangulares de un punto, y dicho punto se asocia al número complejo $a + bi$, entonces la adición y sustracción de números complejos corresponde a la

adición y sustracción de vectores, es decir, de segmentos dirigidos desde el origen a los puntos de coordenadas (a, b) y (c, d) , puesto que para la adición de vectores se suman sus correspondientes coordenadas. En cuanto al significado geométrico del producto en el plano de los números complejos, resulta más fácil operar con la longitud ρ del vector que va desde el origen del sistema de coordenadas al punto (x, y) (esta longitud se llama *módulo* del número complejo $z = x + iy$) y con el ángulo ϕ que el vector forma con el eje Ox (este ángulo se llama *argumento* del número com-

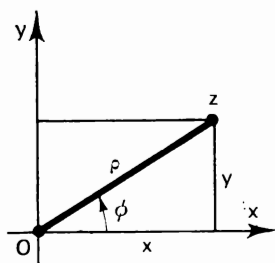


FIG. 1

plejo $z = x + iy$); en otras palabras, se trata de operar con las coordenadas polares ρ y ϕ (Fig. 1) en lugar de las coordenadas cartesianas x e y . Entonces $x = \rho \cos \phi$, $y = \rho \sin \phi$, y por tanto el número complejo $x + iy$ puede escribirse como

$$x + iy = \rho(\cos \phi + i \sin \phi).$$

Si

$$a + bi = \rho_1(\cos \phi_1 + i \sin \phi_1), \quad c + di = \rho_2(\cos \phi_2 + i \sin \phi_2),$$

entonces

$$ac - bd = \rho_1 \rho_2 (\cos \phi_1 \cos \phi_2 - \sin \phi_1 \sin \phi_2) = \rho_1 \rho_2 \cos(\phi_1 + \phi_2),$$

$$bc + ad = \rho_1 \rho_2 (\sin \phi_1 \cos \phi_2 + \cos \phi_1 \sin \phi_2) = \rho_1 \rho_2 \sin(\phi_1 + \phi_2),$$

lo que demuestra que para multiplicar dos números complejos se multiplican sus módulos ϱ_1 y ϱ_2 , y se suman sus argumentos ϕ_1 y ϕ_2 . En la división, puesto que es operación inversa de la multiplicación, se divide un módulo por el otro, y se restan los argumentos:

$$\begin{aligned}\varrho_1(\cos \phi_1 + i \operatorname{sen} \phi_1) \varrho_2(\cos \phi_2 + i \operatorname{sen} \phi_2) &= \\ &= \varrho_1 \varrho_2 [\cos (\phi_1 + \phi_2) + i \operatorname{sen} (\phi_1 + \phi_2)]\end{aligned}$$

y

$$\frac{\varrho_1(\cos \phi_1 + i \operatorname{sen} \phi_1)}{\varrho_2(\cos \phi_2 + i \operatorname{sen} \phi_2)} = \frac{\varrho_1}{\varrho_2} [\cos (\phi_1 - \phi_2) + i \operatorname{sen} (\phi_1 - \phi_2)].$$

Para elevar un número complejo a una potencia de exponente entero positivo n , se eleva el módulo a la potencia n -ésima y se multiplica el argumento por n

$$[\varrho(\cos \phi + i \operatorname{sen} \phi)]^n = \varrho^n (\cos n\phi + i \operatorname{sen} n\phi).$$

Recíprocamente, la raíz n -ésima es:

$$\sqrt[n]{\varrho(\cos \phi + i \operatorname{sen} \phi)} = \sqrt[n]{\varrho} \left(\cos \frac{\phi}{n} + i \operatorname{sen} \frac{\phi}{n} \right)$$

Sin embargo, al calcular raíces nos enfrentamos con una situación especial. Sea n un exponente entero positivo. Entonces

$$\sqrt[n]{\varrho(\cos \phi + i \operatorname{sen} \phi)}$$

es igual al número

$$\sqrt[n]{\varrho} \left(\cos \frac{\phi}{n} + i \operatorname{sen} \frac{\phi}{n} \right),$$

puesto que al elevar este número a la potencia n -ésima se obtiene el radicando.

Pero este es sólo uno de los valores de la raíz. Lo interesante es que el número complejo

$$\sqrt[n]{\varrho} \left[\cos \left(\frac{\phi}{n} + \frac{2k\pi}{n} \right) + i \operatorname{sen} \left(\frac{\phi}{n} + \frac{2k\pi}{n} \right) \right],$$

donde k es cualquiera de los números $1, 2, \dots, n-1$, es también raíz n -ésima del número

$$\varrho(\cos \phi + i \operatorname{sen} \phi).$$

En efecto, según la regla de potenciación, si elevamos este número a la potencia n -ésima se obtiene el número

$$\begin{aligned} (\sqrt[n]{\varrho})^n \left[\cos n \left(\frac{\phi}{n} + \frac{2k\pi}{n} \right) + i \operatorname{sen} n \left(\frac{\phi}{n} + \frac{2k\pi}{n} \right) \right] = \\ = \varrho [\cos (\phi + 2k\pi) + i \operatorname{sen} (\phi + 2k\pi)], \end{aligned}$$

donde el sumando $2k\pi$, por las propiedades del seno y el coseno, puede suprimirse, puesto que al hacerlo no varía ni el seno ni el coseno. Así, la potencia n -ésima de este número es también

$$\varrho(\cos \phi + i \operatorname{sen} \phi),$$

es decir, este número es

$$\sqrt[n]{\varrho}(\cos \phi + i \operatorname{sen} \phi).$$

Es fácil ver que no hay ningún otro número complejo, además de los n que corresponden a $k = 0, 1, 2, \dots, n-1$, que sea raíz n -ésima de

$$\varrho(\cos \phi + i \operatorname{sen} \phi).$$

Geométricamente, la extracción de las raíces n -ésimas puede describirse como sigue. Los puntos del plano complejo que corresponden a los valores de la $\sqrt[n]{\varrho}$ del número $\varrho(\cos \phi + i \operatorname{sen} \phi)$ se encuentran sobre los vértices de un polígono regular de n lados inscrito en una circunferencia de centro el origen y radio $\sqrt[n]{\varrho}$ y dispuestos de modo que uno de los vértices de este polígono de n lados tenga argumento ϕ/n (Fig. 2).

Hagamos ahora la siguiente observación. Si

$$f(z) = z^n + c_1 z^{n-1} + \dots + c_{n-1} z + c_n$$

es un polinomio en z de coeficientes reales o complejos $c_1, c_2, \dots, c_n$ y variamos z con continuidad, es decir, movemos continuamente el punto $z = x + iy$ en el plano complejo, entonces el punto $Z = X + iY = f(z)$ también se moverá con continuidad en el plano complejo. Esto se deduce del hecho de que, si sustituimos en $f(z)$ los

valores de $z = x + iy$, $c_1 = a_1 + b_1i$, $c_2 = a_2 + b_2i, \dots, c_n = a_n + b_ni$, y realizamos todos los cálculos, encontramos que

$$f(z) = X + iY,$$

donde

$$X = P(x, y), \quad Y = Q(x, y)$$

son polinomios de grado n en x e y con coeficientes reales expresados en función de a_i y b_i . Bajo cambios continuos de x e y , estos polinomios también variarán con continuidad.

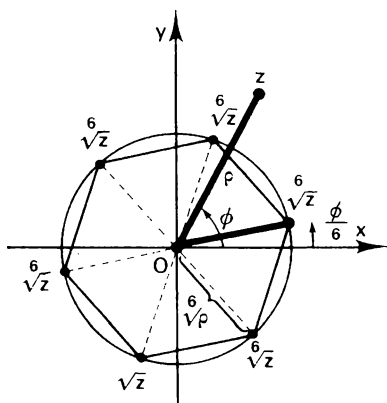


FIG. 2

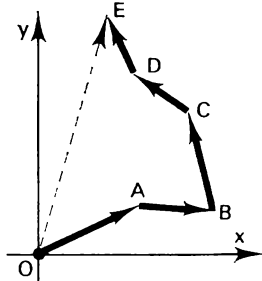


FIG. 3

Señalemos también que, como el módulo $\rho = |f(z)|$ es igual a $\sqrt{X^2 + Y^2}$, en un movimiento continuo del punto z en el plano complejo, el módulo $|f(z)|$ también variará con continuidad. En otras palabras, si el punto z está suficientemente próximo al punto α , entonces la diferencia $|f(z)| - |f(\alpha)|$ de los valores absolutos se puede hacer más pequeña que cualquier número positivo fijado de antemano.

Observemos también que el módulo de una suma de varios números complejos es siempre menor o igual que la suma de los módulos de estos números, lo que es equivalente a decir que el segmento rectilíneo OE (Fig. 3) es más corto o igual que la línea poligonal

$OABCDE$, siendo igual a ella si, y sólo si, todos sus segmentos se encuentran sobre una misma recta y con igual dirección.

Recordemos finalmente que decir que «un número complejo es igual a cero» es lo mismo que decir que «su módulo es igual a cero», puesto que el módulo ϱ de un número complejo es la distancia entre el origen y el punto correspondiente.

Apliquemos ahora la teoría de los números complejos a la demostración del teorema fundamental del álgebra, aunque conviene observar que la importancia de la teoría de los números complejos va mucho más allá de los límites del álgebra. Son muchas las partes de la matemática, aparte del álgebra, que no se podrían desarrollar sin la ayuda de esta clase de números. En muchas aplicaciones —por ejemplo, en la teoría de las corrientes alternas— son numerosos los problemas que se resuelven con sencillez por medio de los números complejos. Pero lo más importante es la aplicación de los números complejos, o, con más precisión, de la teoría de funciones de una variable compleja, a la teoría de ciertas funciones especiales de dos variables reales llamadas *armónicas*. Por medio de estas funciones se pueden resolver importantes problemas en la teoría del vuelo aéreo, de la conducción del calor en una placa, de los campos eléctricos planos y de la elasticidad. Un célebre teorema sobre la fuerza ascensional en un ala de avión la obtuvo el fundador de la aerodinámica contemporánea, N. E. Zúkovski, estudiando funciones de variable compleja ⁸.

Pasemos ahora a la demostración del teorema fundamental del álgebra.

Teorema

Cualquier polinomio

$$f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n,$$

cuyos coeficientes

$$a_0, a_1, \dots, a_{n-1}, a_n$$

son números reales o complejos dados, tiene al menos una raíz real o compleja.

Supondremos que el polinomio dado es de grado n , es decir, que $a_0 \neq 0$.

⁸ Ver capítulo 9.

La superficie del módulo de un polinomio

Consideremos el problema geoméricamente. Sobre cada punto z del plano complejo levantamos un segmento perpendicular de altura t , igual en longitud al módulo $|f(z)|$ del polinomio $f(z)$ en el punto z . Los extremos de estos segmentos definen una superficie M , que puede llamarse superficie del módulo del polinomio $f(z)$. Vemos que esta superficie: 1) nunca desciende por debajo del plano complejo, puesto que el módulo de cualquier número complejo —en este caso, el número $f(z)$ — es no negativo; 2) para cualquier punto dado del plano complejo, la superficie tiene uno y sólo un punto que, o bien se encuentra verticalmente por encima de él, o coincide con él, es decir, la superficie M se extiende como una lámina por encima del plano complejo y en algunos puntos puede incluso tocar el propio plano; 3) la superficie es continua, en el sentido de que un cambio continuo en la posición del punto z sobre el plano complejo produce un cambio continuo en el valor de $t = |f(z)|$, es decir, en la altura t de los puntos de la superficie. (Esto se demostró en la última subsección.)

El teorema fundamental del álgebra consiste en demostrar que la superficie M toca al plano complejo al menos en un punto, sin permanecer en todo punto a una distancia positiva por encima de él.

Crecimiento del módulo de un polinomio con la distancia al origen

Demostremos que por muy grande que sea un número positivo G dado, se puede encontrar un radio R tal que, para todos los puntos z del plano complejo situados fuera del círculo de radio R y centro el origen, la altura t de los puntos de la superficie M sobre el plano complejo es mayor que G .

Escribamos para ello el polinomio $f(z)$ en la forma

$$a_0 z^n \left[1 + \left(\frac{a_1}{a_0 z} + \frac{a_2}{a_0 z^2} + \cdots + \frac{a_n}{a_0 z^n} \right) \right].$$

El módulo de la expresión

$$\left(\frac{a_1}{a_0 z} + \frac{a_2}{a_0 z^2} + \cdots + \frac{a_n}{a_0 z^n} \right)$$

es no mayor que la suma de los módulos de los sumandos

$$\left| \frac{a_1}{a_0 z} \right| + \left| \frac{a_2}{a_0 z^2} \right| + \cdots + \left| \frac{a_n}{a_0 z^n} \right|,$$

y al aumentar el módulo de z , cada uno de estos sumandos decrece, por lo que la suma también disminuye. Por tanto, para todo z cuyo módulo sea mayor que cierto número R' , el módulo de la expresión entre paréntesis es más pequeño, por ejemplo, que $1/2$.

Pero entonces, para todo z la expresión

$$\Omega = \left[1 + \left(\frac{a_1}{a_0 z} + \frac{a_2}{a_0 z^2} + \cdots + \frac{a_n}{a_0 z^n} \right) \right]$$

tendrá un módulo mayor que $1/2$. El módulo del primer factor $a_0 z^n$ es igual a $|a_0| \cdot |z|^n$, por lo cual crecerá al aumentar el módulo de z ; además crece ilimitadamente. Por tanto, por grande que sea un número positivo G dado, existe un número positivo R tal que, para todo z de módulo mayor que R , $|f(z)| = |a_0| \cdot |z|^n \cdot |\Omega|$ es mayor que G .

La existencia de mínimos de la superficie M

Diremos que en un punto a del plano complejo la superficie tiene un mínimo si el valor de la altura t del punto de la superficie M en este punto a es menor, o igual, que los valores en todos los puntos de algún entorno del punto a , es decir, en todos los puntos de algún círculo, suficientemente pequeño, de centro el punto a .

Sea g la altura t del punto de la superficie M correspondiente al origen (el punto $z = 0$ del plano complejo), es decir $|f(0)| = g$. Tomemos $G > g$. Todas las alturas t de los puntos de la superficie M son no negativas y varían continuamente al desplazarse continuamente el punto z en el plano complejo. La superficie M tiene altura $t > G$ fuera de un círculo de centro el origen, radio R y altura $t = g < G$ en el centro del círculo. D'Alembert supuso, como consecuencia obvia, que en alguna parte del interior del círculo R existe un punto en el que la altura es mínima; con más precisión, donde el valor de t es menor o igual que sus valores en todos los restantes puntos del círculo R , es decir: la superficie M tiene al menos un mínimo.

La demostración rigurosa de la existencia de ese mínimo se basa

en el siguiente *axioma de continuidad* del conjunto de los números reales.

Dadas dos sucesiones de números reales: $a_1 \leq a_2 \leq \dots \leq a_n \leq \dots$ y $b_1 \geq b_2 \geq \dots \geq b_n \geq \dots$, tales que $b_n > a_n$ para todo n , y $b_n - a_n \rightarrow 0$ cuando $n \rightarrow \infty$, entonces existe un número real c , y sólo uno, tal que $a_n \leq c \leq b_n$ para todo n .

Geométricamente, esta propiedad de continuidad significa que, si se da sobre la recta una sucesión de intervalos $[a_n, b_n]$ (Fig. 4) tal que todo intervalo esté contenido en el precedente y tal que las lon-

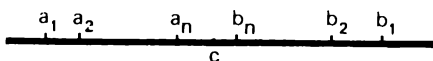


FIG. 4

gitudes de los intervalos se hagan arbitrariamente pequeñas, existe un punto c que pertenece a todos los intervalos de la sucesión. En otras palabras, los intervalos se «contraen» a un punto, no a un «lugar vacío».

Puesto que la longitud del segmento $[a_n, b_n]$ tiende a cero cuando n tiende a infinito, el punto c es único. De la propiedad de continuidad para el conjunto de todos los puntos de la recta real se deduce inmediatamente la propiedad de continuidad para los números complejos, es decir, para los puntos del plano. A continuación damos una formulación geométrica de esta propiedad.

Si se da en el plano una sucesión de rectángulos $\Delta_1, \Delta_2, \dots, \Delta_n, \dots$ de lados paralelos a los ejes de coordenadas y tal que todo rectángulo esté contenido en el anterior y la longitud de sus diagonales decrezca indefinidamente, existe entonces un punto, y sólo uno, contenido en todos los rectángulos de la sucesión. Esta propiedad de continuidad del plano se sigue directamente de la propiedad de continuidad de la recta. Para la demostración es suficiente proyectar los rectángulos sobre los ejes de coordenadas.

Ahora es fácil establecer el llamado teorema de Bolzano-Weierstrass.

Dada una sucesión infinita de puntos $z_1, z_2, \dots, z_n, \dots$ dentro de un rectángulo, existe entonces, en el interior o en la frontera del rectángulo, un punto z_0 tal que en cualquier entorno arbitrariamente pequeño de z_0 , es decir, en el interior de todo círculo de centro z_0 por pequeño que sea, hay un número infinito de puntos de la sucesión $z_1, z_2, \dots, z_n, \dots$.

Para la demostración notaremos por Δ_1 el rectángulo dado. Lo dividimos en cuatro partes iguales mediante rectas paralelas a los ejes de coordenadas. Al menos una de las partes contiene necesariamente infinitos puntos de la ecuación dada. Esta parte se notará por Δ_2 . Dividimos otra vez el rectángulo Δ_2 en cuatro partes iguales y seleccionamos entre ellas una Δ_3 que contenga un número infinito de puntos de la sucesión dada, etc.

De este modo se obtiene una sucesión de rectángulos encajados $\Delta_1, \Delta_2, \Delta_3, \dots$, cuyas diagonales decrecen indefinidamente. Por la propiedad de continuidad podemos encontrar un punto z_0 contenido en todos estos rectángulos. Entonces z_0 es el punto buscado. Porque, por muy pequeño que tomemos el entorno de z_0 , los rectángulos de la sucesión $\Delta_1, \Delta_2, \Delta_3, \dots$, de uno de ellos en adelante, estarán dentro del entorno en cuanto sus diagonales se hagan más pequeñas que el radio del entorno, y cualquiera de los rectángulos contiene un número infinito de puntos de la sucesión $z_1, z_2, \dots, z_n, \dots$. Queda así demostrado el teorema de Bolzano-Weierstrass.

Ahora es ya muy fácil demostrar el teorema del módulo mínimo $|f(z)|$ de un polinomio. Como antes, pongamos $|f(0)| = g$, sea G un número mayor que g , y sea R tal que para $z > R$ se tenga $|f(z)| > G$.

Si $g = 0$, es decir, $f(0) = 0$, entonces el módulo $|f(z)|$ del polinomio tiene un mínimo en el punto 0, puesto que en todos los puntos es ≥ 0 .

Si $g > 0$ y $|f(z)| \geq g$ para todo z , entonces $|f(z)|$ tiene también un mínimo en el punto 0. Supongamos que $g > 0$ y que existen puntos z para los cuales $|f(z)| < g$; entonces en la sucesión de números

$$0, \frac{g}{n}, \frac{2g}{n}, \dots, \frac{ng}{n} = g \quad (*)$$

encontramos el mayor $c_n = (i/n)g$ tal que para todo z sea $|f(z)| \geq c_n$. Para el siguiente número $c'_n = [(i+1)/n]g$ la sucesión (*) contiene al menos un punto z_n tal que $|f(z_n)| < c'_n$.

Hagamos tender n a infinito. Para todo n tenemos $|z_n| \leq R$, puesto que si $|z_n| > R$, entonces $|f(z_n)|$ sería mayor que G y mayor por tanto que g .

Así, todos los puntos z_n se encuentran dentro de un rectángulo de lado $2R$ y de centro el origen. Es posible que algunos de estos puntos coincidan.

Por el teorema de Bolzano-Weierstrass existe un punto z_0 tal que todo entorno de z_0 contiene infinitos puntos de la sucesión $z_1, z_2, \dots, z_n, \dots$.

Demostremos que el punto z_0 es el mínimo de $|f(z)|$ que buscábamos.

En efecto, en cualquier punto z se tiene

$$\begin{aligned} |f(z)| > c_n &= c'_n - \frac{g}{n} > |f(z_n)| - \frac{g}{n} = \\ &= |f(z_0)| + \left[|f(z_n)| - |f(z_0)| \right] - \frac{g}{n}. \end{aligned}$$

Esta desigualdad es válida para cualquier n . Si tomamos para n una sucesión de valores para los cuales z_n se aproxima indefinidamente a z_0 , entonces, teniendo en cuenta la continuidad de $|f(z)|$, la diferencia $|f(z_n)| - |f(z_0)|$ se hace tan pequeña como se quiera, y lo mismo ocurre con g/n .

Por tanto, $|f(z)| \geq |f(z_0)|$, es decir, $|f(z)|$ tiene realmente un mínimo en el punto z_0 .

Lema de d'Alembert

En vista del hecho de que todas las alturas t de los puntos sobre la superficie M de los módulos son no negativas, es claro que cualquier raíz del polinomio $f(z)$, es decir, cualquier punto z del plano complejo donde el polinomio $f(z)$ (y por consiguiente también su módulo $|f(z)|$) sea igual a cero, corresponde a un mínimo de la superficie de los módulos M . El recíproco, como demostró d'Alembert, también es cierto: en cualquier mínimo la superficie M baja hasta el plano complejo, y por consiguiente ese punto es una raíz del polinomio $f(z)$. En otras palabras, en cualquier punto en el que la altura t sea positiva y no nula, no hay mínimo de la superficie M . Esto se sigue del llamado lema de d'Alembert:

Si a es un número complejo dado tal que $f(a) \neq 0$, entonces siempre se puede encontrar un número complejo h de módulo arbitrariamente pequeño, tal que $|f(a+h)| < |f(a)|$.

Demostración

Consideremos el polinomio

$$f(a+h) = a_0(a+h)^n + a_1(a+h)^{n-1} + \cdots + a_{n-1}(a+h) + a_n$$

en dos indeterminadas a y h y ordenémoslo en potencias crecientes de h . En este polinomio hay un término que no contiene a h , a saber:

$$a_0 a^n + a_1 a^{n-1} + \cdots + a_{n-1} a + a_n = f(a) \neq 0,$$

puesto que $f(a) \neq 0$ por hipótesis. Habrá también un término en h^n , el $a_0 h^n$, puesto que se supuso que $a_0 \neq 0$. En cuanto a los términos con potencias intermedias de h , algunos de ellos, y en algunos casos todos, puede que no aparezcan. Sea m la potencia más baja de h que aparece en este polinomio, donde $1 \leq m \leq n$, es decir, esta expresión tendrá la forma

$$f(a+h) = f(a) + Ah^m + Bh^{m+1} + Ch^{m+2} + \cdots + a_0 h^n.$$

Escribámosla así:

$$f(a+h) = f(a) + Ah^m + Ah^m \left(\frac{B}{A} h + \frac{C}{A} h^2 + \cdots + \frac{a_0}{A} h^{n-m} \right),$$

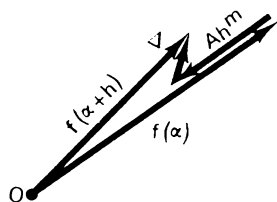


FIG. 5

donde $A \neq 0$, y B , C , etc., pueden ser o no iguales a cero.

Después de esta preparación, la demostración del lema de d'Alembert continúa como sigue. Para h es suficiente tomar un número complejo de módulo suficientemente pequeño para que la longitud del vector Ah^m sea más pequeña que la longitud del vector $f(a)$, y de argumento tal que la dirección del vector Ah^m sea opuesta a la dirección del vector $f(a)$. Entonces el vector $f(a) + Ah^m$ será más pequeño que el vector $f(a)$. Pero si tomamos el módulo h suficientemente pequeño, el módulo de la expresión

$$\left(\frac{B}{A} h + \frac{C}{A} h^2 + \cdots + \frac{a_0}{A} h^{n-m} \right)$$

puede hacerse arbitrariamente pequeño, por ejemplo, menor que 1; por tanto, el vector

$$\Delta = Ah^m \left(\frac{B}{A} h + \frac{C}{A} h^2 + \cdots + \frac{a_0}{A} h^{n-m} \right)$$

es más pequeño que el vector Ah^m , y el vector $f(a+h) = f(a) + Ah^m + \Delta$, como se ve en la figura 5, es también más pequeño que el vector $f(a)$, incluso si la dirección del vector Δ es opuesta a la del vector Ah^m .

Los detalles de esta demostración son los siguientes:

1. Puesto que en el producto se suman los argumentos de los factores, tenemos que tomar el argumento de h tal que

$$\arg A + m \cdot \arg h = \arg f(a) + 180^\circ,$$

es decir, es necesario tomar

$$\arg h = \frac{\arg f(a) - \arg A + 180^\circ}{m}$$

2. El módulo de

$$\left(\frac{B}{A} h + \frac{C}{A} h^2 + \cdots + \frac{a_0}{A} h^{n-m} \right)$$

es no mayor que la suma de los módulos de sus sumandos:

$$T = \left| \frac{B}{A} h \right| + \left| \frac{C}{A} h^2 \right| + \cdots + \left| \frac{a_0}{A} h^{n-m} \right|;$$

además, al decrecer el módulo de h , cada uno de los sumandos de esta suma puede hacerse arbitrariamente pequeño, y lo mismo puede hacerse con la suma. Por tanto, si h es un número complejo con el argumento anterior, y h_0 es (en módulo) tal que si h tiene un módulo más pequeño que h_0 satisface las dos condiciones $|Ah^m| < |f(a)|$ y $T < 1$, entonces, para un tal h , tendremos $|f(a+h)| < |f(a)|$, lo que demuestra el lema de d'Alembert.

Del lema de d'Alembert se sigue inmediatamente que todo mínimo de la superficie de módulos M del polinomio $f(z)$ da una raíz de este polinomio. En efecto, si en el punto a se cumple $f(a) \neq 0$, entonces, en virtud de lema de d'Alembert para puntos arbitrariamente próximos $a+h$, tendríamos $|f(a+h)| < |f(a)|$, es decir, no existiría un círculo de centro a tal que en todos sus puntos el módulo de $f(z)$ fuese no menor que el módulo de $|f(a)|$, y por tanto en el punto a no tendríamos un mínimo del módulo de $f(z)$. Con esto queda demostrado el teorema fundamental del álgebra.

Forma general de la superficie de módulos M

La superficie de módulos M del polinomio $f(z)$ se encuentra por encima del plano complejo z y tiene la forma dibujada en la figura 6. Puede demostrarse que, para grandes alturas t , la superficie M difiere

muy poco de la superficie obtenida por revolución de la parábola de grado n , $t = |a_0| x^n$, alrededor del eje Ot . Para valores de t pequeños la superficie M tiene mínimos, en número igual al de raíces distintas de la ecuación $f(z) = 0$. En todos estos mínimos la superficie M toca al plano complejo z .

§4. Estudio de la distribución de las raíces de un polinomio sobre el plano complejo

Existe una serie de problemas prácticos importantes que se hallan relacionados con la siguiente cuestión: sin resolver una ecuación dada, obtener información acerca de la distribución de sus raíces en el

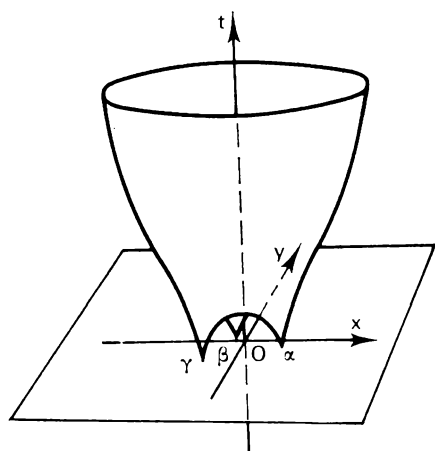


FIG. 6

plano complejo. El primero de tales problemas fue, históricamente, el de determinar el número de raíces reales de una ecuación; esto es, dada una ecuación de coeficientes reales, determinar (por algún criterio, que dependerá de sus coeficientes, y sin resolver la ecuación) si tiene raíces reales y, en caso afirmativo, cuántas; o cuántas raíces positivas y cuántas negativas tiene; o cuántas raíces reales se encuentran entre los límites a y b .

Derivadas de un polinomio

En esta sección será de importancia esencial la derivada de un polinomio. La definición de derivada de una función ya se dio en el cap. 2.

Para el polinomio $a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n$ su derivada viene dada, como es sabido, por el polinomio

$$na_0x^{n-1} + (n-1)a_1x^{n-2} + \dots + a_{n-1}.$$

En el capítulo 2 definimos el concepto de derivada únicamente para funciones de variable real. En álgebra es necesario suponer que

la variable toma valores complejos arbitrarios e introducir polinomios con coeficientes complejos.

Sin embargo, podemos mantener la antigua definición de derivada como límite del cociente del incremento de la función y el incremento de la variable independiente cuando este último tiende a 0. La fórmula para calcular la derivada de un polinomio de coeficientes complejos y las reglas básicas para la derivada de la suma, producto y potencia son las mismas que antes ⁹.

Raíces simples y múltiples de un polinomio

En el §2 de este capítulo se estableció que, si el número a es raíz del polinomio $f(x)$, entonces $f(x)$ es divisible exactamente por $x - a$. Si $f(x)$ no es divisible por $(x - a)^2$, entonces se dice que el número a es raíz simple del polinomio $f(x)$. En general, si el polinomio $f(x)$ es divisible por $(x - a)^k$ pero no por $(x - a)^{k+1}$, entonces se dice que el número a es una raíz de multiplicidad k .

Una raíz a de multiplicidad k se considera a menudo como k raíces distintas. Ello es así porque el factor $(x - a)^k$, presente en la factorización de $f(x)$ en factores lineales, es el producto de k factores, todos iguales a $(x - a)$.

En virtud del hecho de que todo polinomio de grado n puede ser factorizado en producto de n factores lineales, el número de raíces de un polinomio es igual a su grado, teniendo en cuenta la multiplicidad de cada raíz.

Se verifican los siguientes teoremas:

1. Una raíz simple de un polinomio no es raíz de su derivada.
2. Una raíz múltiple de un polinomio es raíz (con multiplicidad inferior en una unidad) de su derivada.

En efecto, sea $f(x) = (x - a)^k f_1(x)$, donde $f_1(x)$ no es divisible por $(x - a)$, es decir, $f_1(a) \neq 0$. Entonces

$$\begin{aligned} f'(x) &= k(x - a)^{k-1} f_1(x) + (x - a)^k f_1'(x) = \\ &= (x - a)^{k-1} [k f_1(x) + (x - a) f_1'(x)] = (x - a)^{k-1} F(x). \end{aligned}$$

El polinomio $F(x) = k f_1(x) + (x - a) f_1'(x)$ no es divisible por $(x - a)$, puesto que $F(a) = k f_1(a) \neq 0$.

Por tanto para $k = 1$, $f'(x)$ no es divisible por $(x - a)$, y para $k > 1$, es divisible por $(x - a)^{k-1}$ pero no por $(x - a)^k$. Con esto quedan demostrados ambos teoremas.

⁹ Ver capítulo 9.

Teorema de Rolle y algunas de sus consecuencias

Según el conocido teorema de Rolle¹⁰, si los números reales a y b son raíces de un polinomio de coeficientes reales, existe un número c entre a y b que es raíz de la derivada.

Del teorema de Rolle se deducen algunos teoremas interesantes:

1. Si todas las raíces del polinomio $f(x) = a_0x^n + \dots + a_n$ son reales, todas las raíces de su derivada son también reales. Además, entre dos raíces consecutivas de $f(x)$ sólo existe una raíz de $f'(x)$, y ésta es simple. En efecto, sean $x_1 < x_2 < \dots < x_k$ las raíces de $f(x)$, con multiplicidades $m_1, m_2, \dots, m_k$, respectivamente. Es obvio que $m_1 + m_2 + \dots + m_n = n$.

Por el teorema anterior sobre las raíces múltiples, la derivada $f'(x)$ tendrá raíces $x_1, x_2, \dots, x_k$ de multiplicidades $m_1 - 1, m_2 - 1, \dots, m_k - 1$, y por el teorema de Rolle habrá al menos una raíz $y_1, y_2, \dots, y_{k-1}$ en el interior de cada uno de los intervalos $(x_1, x_2), (x_2, x_3), \dots, (x_{k-1}, x_k)$ entre dos raíces consecutivas de $f(x)$. Así, el número de raíces reales de $f'(x)$ es igual (considerando las multiplicidades) a $(m_1 - 1) + (m_2 - 1) + \dots + (m_k - 1) + k - 1 = n - 1$ por lo menos. Pero $f'(x)$, como polinomio de grado $(n - 1)$, posee (teniendo en cuenta las multiplicidades) $n - 1$ raíces. Por tanto, todas las raíces de $f'(x)$ son reales, $y_1, y_2, \dots, y_{k-1}$ son raíces simples, y no existen otras raíces del polinomio $f'(x)$ distintas de $x_1, x_2, \dots, x_k$ e $y_1, y_2, \dots, y_{k-1}$.

2. Si todas las raíces de un polinomio $f(x)$ son reales y de ellas p son positivas, entonces $f'(x)$ tiene p ó $p - 1$ raíces positivas.

En efecto, sean $x_1 < x_2 < \dots < x_k$ todas las raíces positivas del polinomio $f(x)$ con multiplicidades $m_1, m_2, \dots, m_k$, respectivamente. Entonces $m_1 + m_2 + \dots + m_k = p$. La derivada $f'(x)$ tendrá las siguientes raíces positivas: $x_1, x_2, \dots, x_k$ de multiplicidades $m_1 - 1, m_2 - 1, \dots, m_k - 1$, por un lado; las raíces simples $y_1, y_2, \dots, y_{k-1}$ en los intervalos $(x_1, x_2), \dots, (x_{k-1}, x_k)$, por otro; y puede también tener una raíz simple y_0 en el intervalo (x_0, x_1) , donde x_0 es la mayor raíz no positiva de $f(x)$. Por tanto, el número de raíces positivas es igual a $(m_1 - 1) + \dots + (m_k - 1) + K - 1 = p - 1$, o sea, $(m_1 - 1) + \dots + (m_k - 1) + (k - 1) + 1 = p$, que es lo que se quería demostrar.

¹⁰ Este teorema es la forma más sencilla del teorema del valor medio, que ya se mencionó en el capítulo 2.

Ley de Descartes de los signos

En su importante libro, publicado en 1637, «Geometría», en el que aparecía la primera presentación de la geometría analítica, Descartes dio, entre otras cosas, el primer teorema algebraico significativo referente a la distribución de las raíces de un polinomio en el plano complejo: la llamada «ley de los signos de Descartes», que puede establecerse como sigue:

Si los coeficientes de una ecuación son reales y se sabe que todas sus raíces son reales, entonces el número de sus raíces positivas (teniendo en cuenta sus multiplicidades) es igual al número de cambios de signo en la sucesión de sus coeficientes. Si también tiene raíces complejas, entonces dicho número es igual al número de cambios de signo, menos un número par.

Explicaremos primero lo que entendemos por número de cambios de signo en la sucesión de los coeficientes de una ecuación. Para obtener este número se escriben todos los coeficientes de la ecuación, en el orden, por ejemplo, de las potencias decrecientes de la incógnita, incluyendo el coeficiente de x^n y el término constante, pero omitiendo los coeficientes iguales a cero, y se consideran todos los pares de números sucesivos así obtenidos de la sucesión. Si en una pareja dada los signos de los números son diferentes, llamamos a esto cambio de signo. Por ejemplo, si la ecuación es

$$x^7 + 3x^5 - 5x^4 - 8x^2 + 7x + 2 = 0$$

entonces la sucesión de sus coeficientes es

$$1, 3, -5, -8, 7, 2,$$

y hay dos cambios de signo.

Pasemos ahora a la demostración de la primera parte del teorema ¹¹.

Sin pérdida de generalidad podemos suponer que el coeficiente a_0 del término de mayor grado del polinomio $f(x) = a_0x^n + \dots + a_n$ es positivo.

En primer lugar, estableceremos si $f(x)$ sólo tiene raíces reales y p de ellas son positivas (teniendo en cuenta las multiplicidades),

¹¹ Se podría dar otra demostración, directa, que no utilizase derivadas, pero sería algo más larga.

entonces $(-1)^p$ es el signo del coeficiente del término de menor grado de $f(x)$ distinto de cero.

En efecto, sea

$$\begin{aligned} f(x) &= a_0 x^n + \cdots + a_k x^{n-k} = \\ &= a_0 x^{n-k} (x - x_1) \cdots (x - x_p) (x - x_{p+1}) \cdots (x - x_{n-k}), \end{aligned}$$

donde $x_1, \dots, x_p$ son las raíces positivas de $f(x)$, y $x_{p+1}, \dots, x_{n-k}$ son las raíces negativas de $f(x)$, habiéndose tenido en cuenta la multiplicidad de cada raíz. Entonces $a_k = a_0 (-1)^p x_1 \cdots x_p (-x_{p+1}) \cdots (-x_{n-k})$ y, puesto que todos los números $a_0, x_1, \dots, x_p, -x_{p+1}, \dots, -x_{n-k}$ son positivos, el signo de a_k es $(-1)^p$ ¹².

La siguiente demostración se basa en el método de inducción completa.

Para los polinomios de primer grado el teorema es trivial. En efecto, un polinomio de primer grado $a_0 x + a_1$ tiene una raíz única $-a_1/a_0$, que es positiva si, y sólo si, a_0 y a_1 tienen signos opuestos.

Supongamos que ya se ha demostrado el teorema para todos los polinomios de grado $(n-1)$, de raíces reales, y con esta suposición lo demostraremos para cualquier polinomio $f(x) = a_0 x^n + \cdots + a_{n-1} x + a_n$ de grado n .

1. $a_n = 0$. Consideremos el polinomio $f_1(x) = a_0 x^{n-1} + \cdots + a_{n-1}$. Las raíces positivas de los polinomios $f(x)$ y $f_1(x)$ son las mismas; el número de cambios de signo en la sucesión de sus coeficientes es también el mismo. Para el polinomio $f_1(x)$ es válida la ley de Descartes; por tanto es válida para el polinomio $f(x)$.

2. $a_n \neq 0$. Consideremos la derivada

$$f'(x) = n a_0 x^{n-1} + (n-1) a_1 x^{n-2} + \cdots + a_{n-1}.$$

Es claro que el número de cambios de signo en la sucesión de los coeficientes del polinomio $f(x)$ es igual al número análogo para la derivada $f'(x)$, si los signos de a_n y el último coeficiente no nulo de la derivada coinciden; y una unidad mayor, si los signos son opuestos.

En virtud de lo que indicamos al comienzo de la demostración, el número de raíces positivas de $f(x)$ y de $f'(x)$ tiene en el primer caso la misma paridad (ambos son pares o impares), y en el segundo caso tiene paridades opuestas. Pero, como se deduce del teorema de Rolle,

¹² Observemos que esta suposición es también correcta en el caso de que algunas de las raíces $f(x)$ sean complejas.

el número de raíces positivas de un polinomio, si todas sus raíces son reales, puede ser o igual al número de raíces positivas de su derivada, o una unidad mayor. Teniendo esto en cuenta, fijémonos en que en el primer caso $f(x)$ tiene el mismo número de raíces positivas que $f'(x)$, y en el segundo caso una más. Para $f'(x)$ la ley de Descartes es válida por la hipótesis de inducción, es decir, el número de raíces positivas de $f'(x)$ es igual al de cambios de signo en la sucesión de sus coeficientes.

Por tanto, en ambos casos el número de raíces positivas de $f(x)$ es igual al de cambios de signo en la sucesión de coeficientes, como queríamos demostrar.

La segunda parte de la ley de Descartes no ofrece mayores complicaciones, y omitiremos su demostración.

Nota 1

La primera parte del teorema de Descartes es particularmente importante, puesto que en muchos problemas prácticos se sabe de modo automático si todas las raíces de una ecuación dada son reales. En este caso puede deducirse rápidamente el número de raíces positivas y negativas, y cuántas raíces nulas tiene la ecuación.

Nota 2

Si en el polinomio ponemos $x = y + a$, donde a es un número real arbitrario dado, es decir, formamos el polinomio $f(y + a)$, entonces las raíces positivas y de este polinomio serán aquellas, y sólo aquellas, que se obtengan a partir de las raíces x del polinomio dado $f(x)$ que sean mayores que a . Por tanto, el número de raíces del polinomio dado $f(x)$ (cuyas raíces son todas reales), comprendidas entre los límites a y b ($b > a$), es igual al número de cambios de signo del polinomio $f(y + a)$ menos el número de cambios de signo del polinomio $f(z + b)$. Sin embargo, si no son reales todas las raíces de $f(x)$, puede entonces demostrarse que este número es igual a dicha diferencia, o a dicha diferencia menos cierto número par. Este es el llamado teorema de Budan.

Teorema de Sturm

La ley de Descartes de los signos, como el teorema de Budan, no dan, sin embargo, solución al problema: si una ecuación de coeficientes reales tiene al menos una raíz real, ¿cuántas raíces reales tiene en total, y cuántas raíces reales tiene en el intervalo cuyos extre-

mos son a y b ? Durante más de dos siglos los matemáticos intentaron resolver estos problemas, pero sin éxito. Descartes, Newton, Sylvester, Fourier y otros muchos trabajaron con denuevo sobre ello, pero no pudieron resolver ni siquiera el primero de estos problemas, hasta que finalmente, en 1835, el matemático francés Sturm sugirió un método que resolvía los tres problemas.

El método de Sturm no es realmente muy complicado, pero su carácter es tal que se le podría buscar por mucho tiempo y no encontrarlo. El mismo Sturm quedó muy satisfecho por su éxito en la resolución de este importantísimo problema algebraico. En sus conferencias, al presentar este resultado solía decir: «He aquí el teorema que lleva mi nombre». Pero hay que decir que Sturm no resolvió este problema por simple casualidad, pues durante muchos años estuvo meditando sobre cuestiones relacionadas con él.

Sea $f(z)$ un polinomio con coeficientes reales y $f_1(z)$ la derivada $f'(z)$. Dividamos el polinomio $f(z)$ por $f_1(z)$ y notemos por $f_2(z)$ el resto de la división, tomándolo con signo opuesto. Luego, dividamos $f_1(z)$ por $f_2(z)$ y notemos el resto, tomado con signo opuesto, por $f_3(z)$, etc.

Puede demostrarse que el último polinomio no nulo $f_s(z)$ de la sucesión así construida será un número constante c .

El teorema de Sturm dice lo siguiente: si $a < b$ son dos números reales que no sean raíces del polinomio $f(z)$, sustituyendo en los polinomios

$$f(z), f_1(z), \dots, f_{s-1}(z), c$$

$z = a$ y $z = b$, obtenemos dos sucesiones de números reales:

$$f(a), f_1(a), f_2(a), \dots, f_{s-1}(a), c, \quad (\text{I})$$

$$f(b), f_1(b), f_2(b), \dots, f_{s-1}(b), c, \quad (\text{II})$$

tales que el número de cambios de signo en la sucesión (I) es mayor o igual al número de cambios de signo en la sucesión (II), y la diferencia entre estos números es exactamente igual al número de raíces reales de $f(z)$ comprendidas entre a y b ; o, dicho con otras palabras, el número de estas raíces es igual a la pérdida de cambios de signo en la sucesión (I) al pasar de a a b .

La demostración del teorema de Sturm no es más difícil que la del teorema de Descartes, pero no se dará aquí.

El teorema de Sturm permite calcular el número de raíces de un

polinomio de coeficientes reales sobre cualquier segmento del eje real. Por tanto, la aplicación del teorema de Sturm a cualquier polinomio proporciona una imagen clara de la distribución de sus raíces sobre el eje real; en particular permite *separar* las raíces, es decir, construir segmentos en cada uno de los cuales sólo esté contenida una raíz del polinomio.

En muchas aplicaciones es igualmente importante la solución del problema análogo para las raíces complejas de un polinomio. Puesto que los números complejos se representan por puntos del plano y no de la recta, es imposible hablar de «segmentos» en los cuales estén contenidas las raíces complejas; en su lugar, habrá que considerar una región, es decir, una porción de plano, elegida de un modo u otro.

Así, con respecto a las raíces complejas se plantea el siguiente problema:

Dado un polinomio $f(z)$ y una región del plano complejo, se pide encontrar el número de raíces del polinomio contenido en esta región.

Supondremos que la región está limitada por un contorno cerrado (Fig. 7) y que sobre dicho contorno el polinomio $f(z)$ no tiene ninguna raíz.

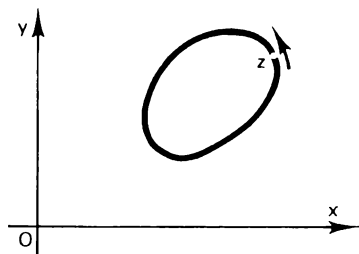


FIG. 7

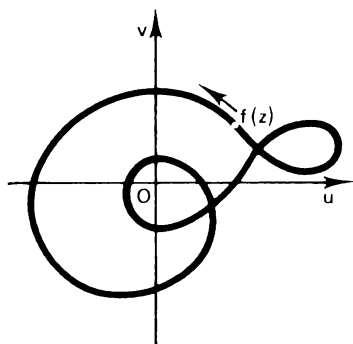


FIG. 8

Supongamos que el punto z describe una vuelta completa a lo largo del contorno de la región, siguiendo la dirección positiva. Todo valor del polinomio viene también representado por puntos del plano. Bajo cambios continuos de z el polinomio $f(z)$ variará también con continuidad. Por tanto, mientras z recorre una vez el contorno

de la región, $f(z)$ describirá una curva cerrada. Esta curva no pasará por el origen del sistema de coordenadas, puesto que $f(z)$, por hipótesis, no se reduce a cero en ninguno de los puntos del contorno (Fig. 8).

La solución al problema mencionado la da el siguiente teorema.

Principio del argumento

El número de raíces del polinomio $f(z)$ dentro de la región limitada por una curva cerrada C es igual al número de vueltas que da el punto $f(z)$ alrededor del origen cuando z recorre una vez el contorno C en dirección positiva.

Para la demostración descompondremos $f(z)$ en factores lineales

$$f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n = a_0(z - z_1)(z - z_2) \cdots (z - z_n).$$

Sabemos que el argumento del producto de varios números complejos es igual a la suma de los argumentos de los factores. Por tanto,

$$\arg f(z) = \arg a_0 + \arg(z - z_1) + \arg(z - z_2) + \cdots + \arg(z - z_n).$$

Notemos por $\Delta \arg f(z)$ el incremento del argumento de $f(z)$, calculado en el supuesto de que z recorra una vez el contorno C . Es claro que $\Delta \arg f(z)$ es 2π multiplicado por el número de vueltas que dé el punto $f(z)$ alrededor del origen.

Evidentemente,

$$\begin{aligned} \Delta \arg f(z) &= \Delta \arg a_0 + \Delta \arg(z - z_1) + \\ &+ \Delta \arg(z - z_2) + \cdots + \Delta \arg(z - z_n). \end{aligned}$$

Es claro que $\Delta \arg a_0 = 0$, puesto que a_0 es constante. Por otra parte, $z - z_1$ viene representado por el vector que va del punto z_1 al punto z . Supongamos que z_1 se halla en el interior de la región. Geométricamente es obvio (Fig. 9) que cuando el punto z recorre el contorno C , el vector $z - z_1$ describe una revolución completa, por lo que $\Delta \arg(z - z_1) = 2\pi$. Supongamos ahora que el punto z_2 está en el exterior de la región. En este caso el vector «oscila» a un lado y a otro, y vuelve a su posición original sin completar ninguna vuelta, por lo que $\Delta \arg(z - z_2) = 0$. Y del mismo modo cabría razonar sobre todas las raíces. Por tanto, $\Delta \arg f(z)$ es igual a 2π multiplicado por el número de raíces de $f(z)$ que se encuentren en el interior de la región. Y el número de raíces de $f(z)$ dentro de la región es igual

al número de vueltas que da el punto $f(z)$ alrededor del origen, que es lo que había que demostrar.

Este teorema permite resolver el problema en cualquier caso particular, y dibujar la curva trazada por el punto $f(z)$ con un grado de precisión arbitrario. Para ello es necesario tomar un conjunto suficientemente denso de puntos z sobre el contorno C , calcular los correspondientes valores de $f(z)$ y unirlos por medio de una curva

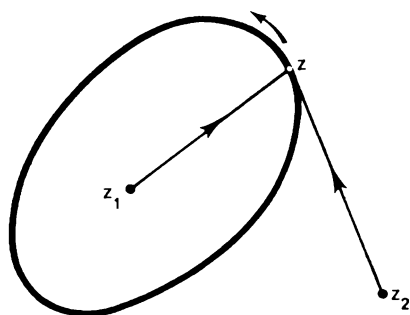


FIG. 9

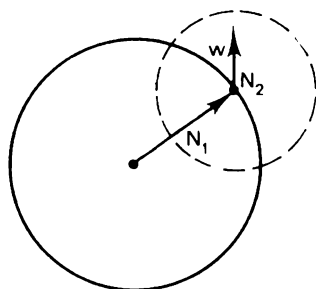


FIG. 10

continua. Sin embargo, en algunos casos se pueden evitar estos tediosos cálculos. Ilustremos uno de los métodos con un ejemplo numérico.

Ejemplo

Encontrar el número de raíces del polinomio $f(z) = z^{11} + 5z^2 - 2$ dentro de un círculo de radio 1 y de centro el origen.

Sobre la circunferencia $|z| = 1$, uno de los tres términos que constituyen el polinomio $f(z)$, el $5z^2$, domina a los demás. En efecto, $|5z^2| = 5$, pero $|z^{11} - 2| \leq |z|^{11} + 2 = 3$. Este hecho permite razonar del siguiente modo. Notemos a $z^{11} + 5z^2 - 2$ por w , $5z^2$ por N_1 , y $z^{11} - 2$ por N_2 . Mientras el punto z recorre una vez la circunferencia unidad, $N_1 = 5z^2$ da dos vueltas alrededor de una circunferencia de radio 5, puesto que $|N_1| = 5$ y $\arg N_1 = 2 \arg z$. El punto w está «sujeto» al punto N_1 por un vector cuya longitud es $|N_2| \leq 3$, es decir, la distancia del punto w al N_1 es siempre menor que la distancia de N_1 al origen del sistema de coordenadas.

Por tanto, el punto w , independientemente de las vueltas que describa alrededor de N_1 (Fig. 10), no puede, sin embargo, girar

«libremente» alrededor del origen, sino que da tantas vueltas alrededor de éste como el punto N_1 , es decir, dos vueltas. Por tanto, el número de raíces de $f(z)$ en el interior de la región es igual a 2.

Problema de Hurwitz

En mecánica, particularmente en la teoría de oscilaciones y control, juegan un papel importante las condiciones que permiten decidir si las partes reales de todas las raíces de un polinomio dado $f(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_n$ (con coeficientes reales) son negativas, es decir, si se encuentran en el semiplano situado a la izquierda del eje imaginario.

Con un razonamiento similar al del principio del argumento se obtiene fácilmente un criterio para resolver este problema.

Supondremos que $a_0 > 0$.

Movamos el punto z (Fig. 11) sobre el eje imaginario y en dirección descendente, es decir, sea $z = iy$ con y variando de $+\infty$ a $-\infty$. Entonces $f(z)$ describe una curva con infinitas ramas. Para este estudio es más conveniente la curva descrita por la función

$$f_1(z) = (i)^{-n} f(z) = a_0 y^n - a_2 y^{n-2} + a_4 y^{n-4} + \dots - \\ - i(a_1 y^{n-1} - a_3 y^{n-3} + \dots) = \phi(y) - i\psi(y),$$

donde

$$\phi(y) = a_0 y^n - a_2 y^{n-2} + \dots; \quad \psi(y) = a_1 y^{n-1} - a_3 y^{n-3} + \dots.$$

Puesto que $\arg i = \pi/2$, se tiene $\arg f_1(z) = -n\pi/2 + \arg f(z)$, y por tanto los incrementos de los argumentos de $f(z)$ y $f_1(z)$ son iguales.

Calculemos el incremento del argumento del punto $f_1(z)$ cuando z se mueve sobre el eje imaginario hacia abajo.

Sea $f(z) = a_0(z - z_1)(z - z_2) \dots (z - z_n)$. Entonces

$$\arg f_1(z) = \arg(a_0 i^{-n}) + \arg(z - z_1) + \arg(z - z_2) + \dots + \arg(z - z_n).$$

Geométricamente es claro que el incremento de $\arg(z - z_k)$ es igual a π si z_k se encuentra en el semiplano de la derecha, y a $-\pi$, si z se encuentra en el semiplano de la izquierda (Fig. 11).

Por tanto, el incremento del argumento de $f_1(z)$ es igual a $\pi(N_1 - N_2)$, donde N_1 es el número de raíces de $f(z)$ en el semiplano de la derecha, y N_2 , el número de raíces en el semiplano de la izquierda. Para que todas las raíces se encuentren en el semiplano de la izquier-

da es necesario y suficiente que el incremento del argumento del punto $f_1(z)$ sea igual a $-\pi n$, es decir, que el punto $f_1(z)$ describa n medias revoluciones en el sentido de las manillas del reloj alrededor del origen (Fig. 12).

Observemos que el punto $f_1(z) = \phi(y) - i\psi(y)$ corta al eje imaginario para aquellos valores de y que son raíces de $\phi(y)$, y al eje real para las raíces de $\psi(y)$. Puesto que $\phi(y)$ no tiene más de n raíces reales y el número de raíces reales de $\psi(y)$ no es mayor que $n - 1$, geométrica-

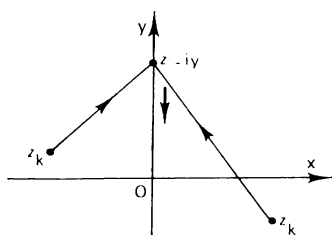


FIG. 11

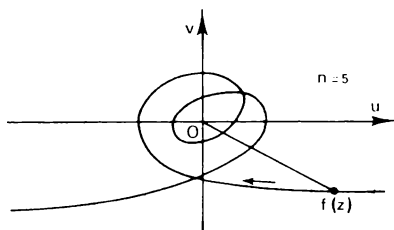


FIG. 12

mente se ve que $f_1(z)$ puede dar n revoluciones completas en el sentido de las manillas del reloj si, y sólo si, la curva viene del cuarto cuadrante y corta consecutivamente al semieje negativo imaginario, al semieje negativo real, al semieje positivo imaginario, al semieje positivo real, etc., por lo que el número total de puntos de intersección con el eje imaginario es n (uno por cada media revolución), y con eje real $n - 1$ (uno menos que el número de semi-revoluciones). Por tanto, el coeficiente a_1 debe ser positivo, y las raíces de los polinomios $\phi(y)$ y $\psi(y)$ deben ser todas reales y alternantes. Esta última afirmación significa que, si $y_1 > y_2 > \dots > y_n$ son las raíces de $\phi(y)$ colocadas en orden decreciente, y $\eta_1 > \eta_2 > \dots > \eta_{n+1}$ son las raíces de $\psi(y)$, entonces $y_1 > \eta_1 > y_2 > \eta_2 > \dots > y_{n-1} > \eta_{n-1} > y_n$.

Así, para que todas las raíces del polinomio $f(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_n$, de coeficientes reales con $a_0 > 0$, se encuentren en el semiplano de la izquierda, es necesario y suficiente que el coeficiente a_1 sea positivo y las raíces de los polinomios $\phi(y) = a_0 y^n - a_2 y^{n-2} + a_4 y^{n-4} - \dots$ y $\psi(y) = a_1 y^{n-1} - a_3 y^{n-3} + \dots$ sean todas reales y alternantes.

Esta condición es equivalente a la conocida condición de Hurwitz de que los determinantes siguientes sean todos positivos:

$$a_1, \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix}, \begin{vmatrix} a_1 & a_0 & a_{-1} \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix}, \dots, \begin{vmatrix} a_1 & a_0 & a_{-1} & \dots & a_{2-n} \\ a_3 & a_2 & a_1 & \dots & a_{4-n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ a_{2n-1} & a_{2n-2} & a_{2n-3} & \dots & a_n \end{vmatrix},$$

donde todas las a_i de índices menores que cero o mayores que n son cero (sobre determinantes, ver capítulo 15, §3).

§5. Cálculo aproximado de raíces

El método de Sturm en combinación con el límite inferior de la diferencia de dos raíces reales distintas permite construir la «separación» de las raíces reales de un polinomio de coeficientes reales, es decir, permite determinar para cada raíz los límites a y b entre los cuales sólo se encuentra dicha raíz. Queda por descubrir un método conveniente para encontrar, en el segmento $a < b$, sucesiones $\alpha_1 < \alpha_2 < \alpha_3 < \dots$ y $\beta_1 > \beta_2 > \beta_3 > \dots$, que converjan tan rápidamente como sea posible a la raíz, aproximándose la primera sucesión por defecto y la segunda por exceso. Cada una de las dos aproximaciones α_k y β_k difiere evidentemente de la raíz en menos de su diferencia $\beta_k - \alpha_k$, puesto que aquélla se encuentra entre ambas aproximaciones. De este modo se puede encontrar una cota superior para el error que se comete al detenernos en un grado de aproximación dado.

Gráfica de un polinomio

Sea

$$f(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n$$

un polinomio de grado n de coeficientes reales.

Consideremos la curva que representa en coordenadas rectangulares la ecuación $y = f(x)$, es decir, la gráfica de este polinomio. Esta curva se llama a veces parábola de orden n . En primer lugar, es claro que para cualquier x real existe una, y sólo una, $y = f(x)$; por tanto, la gráfica f se extiende a todo el eje real. Además, para cam-

bios continuos de x , tanto $f(x)$ como $f'(x)$ varían continuamente, es decir, sin saltos. Por tanto, la gráfica de f es suave y continua. Para valores de x grandes (en valor absoluto), el primer término $a_0 x^n$ supera, en valor absoluto, a la suma de todos los términos restantes, puesto que todos ellos son de grado inferior. De aquí se sigue que si n es par y $a_0 > 0$, entonces la gráfica de f tiende a infinito positivo a la derecha y a la izquierda (y a infinito negativo si $a_0 < 0$); pero si n es impar y $a_0 > 0$, entonces a la derecha tiende a infinito positivo, y a la izquierda, a infinito negativo (y a la inversa si $a_0 < 0$).

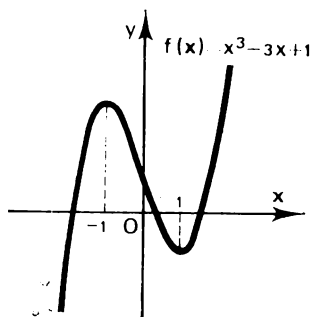


FIG. 13

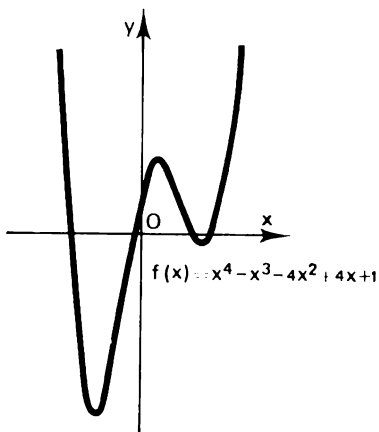


FIG. 14

Los puntos de intersección de la gráfica de f con el eje Ox , es decir, los puntos en los que $y = f(x) = 0$, corresponden a las raíces reales de la ecuación $f(x) = 0$; y de ellos hay n como máximo. En los máximos y mínimos de la gráfica $y = f(x)$, la derivada $f'(x) = 0$; por tanto, el número de máximos y mínimos no es mayor que $n - 1$. Si en alguna sección $f''(x) > 0$, la primera derivada es creciente, y la gráfica será cóncava hacia arriba en dicha sección; si $f''(x) < 0$, entonces la gráfica es cóncava hacia abajo. Puesto que algunas raíces de $f(x) = 0$ pueden ser complejas, el número de máximos y mínimos de la gráfica de f puede ser menor que $n - 1$.

He aquí ejemplos de gráficas de polinomios:

$$f(x) = x^3 - 3x + 1 \text{ (Fig. 13),}$$

$$f(x) = x^4 - x^3 - 4x^2 + 4x + 1 \text{ (Fig. 14).}$$

Después de construida la gráfica de un polinomio es fácil hallar una aproximación de sus raíces. En efecto, las raíces son las abscisas de los puntos de intersección de la gráfica con el eje Ox .

El método del «corto y largo»

Sustituyamos en el polinomio $f(x)$ un número racional entero, por ejemplo 3, y luego sustituyamos 4, 5, 6, ... Si al sustituir 4, 5, 6 obtenemos el mismo signo que para 3, pero para 7 el signo opuesto, entonces es claro que entre 6 y 7 el polinomio $f(x)$ tiene al menos una raíz. Sustituyendo ahora 6; 6, 1; 6, 2; ..., se encontrarán dos números consecutivos de esta sucesión, por ejemplo, 6, 4 y 6, 5, que al sustituirlos dan signos distintos. Así, pues, habrá al menos una raíz entre ellos. Luego sustituimos 6, 4; 6, 41; 6, 42; 6, 43; ..., encontrando así unos límites más próximos aún para la raíz, por ejemplo, 6, 42 y 6, 43, etc. Este es el método del «corto y largo». El método puede simplificarse considerablemente aplicando en cada etapa del cálculo una transformación suplementaria del polinomio, con lo cual sólo será necesario sustituir números enteros, no fracciones, en cada etapa posterior a la primera, y además sólo los números enteros 1, 2, ..., 9. Pero no nos detendremos en esta simplificación.

El método de la tangente y el método de la cuerda

El método de la tangente, llamado método de Newton, y el método de la cuerda, o de interpolación lineal, llamado también método de la posición falsa (*regula falsi*), se emplean por separado o juntos para obtener una estimación del error. Supongamos que entre a y b sólo hay una raíz del polinomio $f(x)$, por lo que $f(a)$ y $f(b)$ tienen signos opuestos, y supongamos también que entre a y b la derivada segunda $f''(x)$ es de signo constante. En este caso la parte de la gráfica de $f(x)$ comprendida entre a y b tiene una de las cuatro formas de la figura 15.

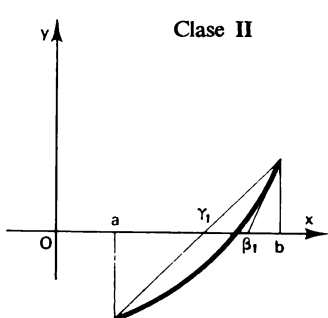
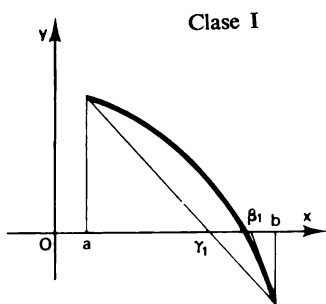
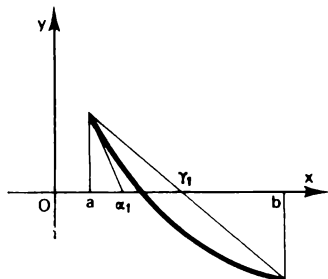
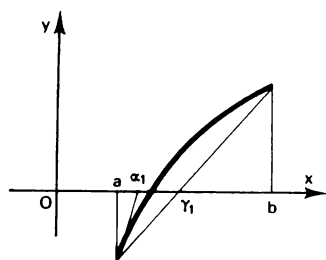
En los casos I y II de la figura 15, la tangente a la gráfica en el punto de abscisa a corta al eje Ox en un punto de abscisa α_1 , situado entre la raíz buscada y a . Si calculamos la abscisa α_1 y consideramos ahora la tangente a la gráfica en el punto de abscisa α_1 , encontramos análogamente un punto α_2 , situado entre el punto α_1 y la raíz buscada, obteniéndose el correspondiente α_3 , y así sucesivamente. De este modo obtendremos aproximaciones por defecto cada vez mejores. Como puede verse en el diagrama, estos valores se aproximan rápidamente a la raíz buscada.

En los casos III y IV es necesario, en cambio, empezar por la

abscisa b , y obtener así puntos $\beta_1, \beta_2, \beta_3, \dots$, es decir, aproximaciones por exceso cada vez más exactas. Es fácil determinar de cual de los cuatro casos se trata, por los signos de $f(a)$, $f(b)$ y $f''(x)$ para $a < x < b$.

Puesto que la ecuación de la tangente a la curva $y = f(x)$ en el punto de abscisa a es

$$y - f(a) = f'(a)(x - a),$$



Clase III

Clase IV

FIG. 15

la abscisa α_1 del punto de intersección con el eje Ox se obtiene a partir de la igualdad

$$0 - f(a) = f'(a)(\alpha_1 - a)$$

esto es

$$\alpha_1 = a - \frac{f(a)}{f'(a)}.$$

Entonces

$$\alpha_2 = \alpha_1 - \frac{f(\alpha_1)}{f'(\alpha_1)}, \quad \alpha_3 = \alpha_2 - \frac{f(\alpha_2)}{f'(\alpha_2)}$$

y así sucesivamente.

Análogamente,

$$\beta_1 = b - \frac{f(b)}{f'(b)}, \quad \beta_2 = \beta_1 - \frac{f(\beta_1)}{f'(\beta_1)}, \quad \beta_3 = \beta_2 - \frac{f(\beta_2)}{f'(\beta_2)}$$

y así sucesivamente.

Este es el método de Newton¹³.

El método de interpolación lineal o de la posición falsa, consiste en lo siguiente. La ecuación de una cuerda, en tanto que recta que pasa por dos puntos dados, tiene la forma

$$\frac{x-a}{b-a} = \frac{y-f(a)}{f(b)-f(a)},$$

y la abscisa γ_1 del punto de su intersección con el eje Ox , obtenida de la ecuación

$$\frac{\gamma_1-a}{b-a} = \frac{0-f(a)}{f(b)-f(a)},$$

es

$$\gamma_1 = -\frac{(b-a)f(a)}{f(b)-f(a)} + a = \frac{af(b)-bf(a)}{f(b)-f(a)}.$$

Tomando este número como nuevo b en los casos I y II y como nuevo a en los casos III y IV, encontramos en los casos I y II

$$\gamma_2 = \frac{af(\gamma_1)-\gamma_1 f(a)}{f(\gamma_1)-f(a)}, \quad \gamma_3 = \frac{af(\gamma_2)-\gamma_2 f(a)}{f(\gamma_2)-f(a)},$$

y así sucesivamente.

¹³ De estas fórmulas se obtiene también una demostración rigurosa de las dos afirmaciones que hemos hecho basándonos en consideraciones sobre los diagramas. Los valores α_n (o β_n) varían monótonamente al crecer n ; por ejemplo, en el caso I crecen y son acotadas, es decir, en virtud del lema de Weierstrass tienden a un límite α . Reemplazando α_n en estas fórmulas por su límite α , obtenemos $\alpha = \alpha - [f(\alpha)/f'(\alpha)]$ de donde $f(\alpha) = 0$, es decir, α es una raíz de f .

En los casos III y IV, tomando γ_1 como nuevo a , encontramos:

$$\gamma_2 = \frac{\gamma_1 f(b) - b f(\gamma_1)}{f(b) - f(\gamma_1)}, \quad \gamma_3 = \frac{\gamma_2 f(b) - b f(\gamma_2)}{f(b) - f(\gamma_2)}$$

y así sucesivamente.

La combinación de estos dos métodos es particularmente útil, puesto que (como puede verse en los diagramas) permite, una vez conocidas las aproximaciones superior e inferior, estimar el error, que es ciertamente menor o igual a la diferencia entre ambas aproximaciones, pues la raíz buscada se encuentra entre una y otra.

Nota

Es importante observar que el hecho de que $f(x)$ sea un polinomio y no otra función de x , no es esencial ni para el método de Newton, ni para el de la interpolación lineal, es decir, ambos métodos pueden aplicarse, en las condiciones antedichas, a ecuaciones transcendentales.

Método de Lobachevski

Uno de los métodos más utilizados para calcular raíces, especialmente raíces complejas, es el método¹⁴ propuesto por N. I. Lobachevski en su libro «Algebra», publicado en 1834. La idea básica de este método procede de Bernoulli.

Observemos, en primer lugar, que, dado un polinomio cuyas raíces son $x_1, x_2, \dots, x_n$, es fácil escribir el polinomio, también de grado n , cuyas raíces son $x_1^2, x_2^2, \dots, x_n^2$, es decir, los cuadrados de las raíces del polinomio dado. En efecto, si $x_1, x_2, \dots, x_n$ son las raíces del polinomio

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_n,$$

éste puede escribirse como

$$(x - x_1)(x - x_2) \dots (x - x_n),$$

y el polinomio

$$x^n - a_1 x^{n-1} + a_2 x^{n-2} - \dots \pm a_n,$$

¹⁴ Este método fue descubierto independientemente por Dandelin (1826), N. I. Lobachevski (1834) y Graeffe (1837).

cuyas raíces son las del polinomio dado tomadas con signo contrario, puede escribirse como

$$(x + x_1)(x + x_2) \cdots (x + x_n).$$

El producto de estos dos polinomios es por consiguiente

$$(x^2 - x_1^2)(x^2 - x_2^2) \cdots (x^2 - x_n^2),$$

que sólo contiene potencias pares de x . Poniendo $x^2 = y$, se obtiene un polinomio de grado n en y :

$$y^n + b_1 y^{n-1} + b_2 y^{n-2} + \cdots + b_n,$$

que puede escribirse

$$(y - x_1^2)(y - x_2^2) \cdots (y - x_n^2),$$

puesto que sus raíces son $x_1^2, x_2^2, \dots, x_n^2$. En vez de multiplicar directamente el polinomio

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_n$$

por el polinomio

$$x^n - a_1 x^{n-1} + a_2 x^{n-2} - \cdots \pm a_n,$$

se puede obtener los coeficientes b_k según el siguiente esquema. Tracemos una línea horizontal, y sobre ella, en la primera fila, se escribe $1, a_1, a_2, \dots, a_n$, y luego, debajo de cada uno de estos coeficientes a_k , se escribe su cuadrado a_k^2 menos dos veces el producto de sus adyacentes

$$-2a_{k-1} a_{k+1},$$

más dos veces el producto de los coeficientes a_{k-2} y a_{k+2}

$$+2a_{k-2} a_{k+2},$$

etcétera, alternando el signo hasta llegar a un coeficiente a partir del cual todos los demás sean cero (a la derecha o a la izquierda de a_k). Los coeficientes b_k se obtienen entonces como suma de las correspondientes columnas de números escritas debajo de la línea.

Obtenidos los coeficientes $1, b_1, b_2, \dots, b_n$ del polinomio cuyas raíces son $1, x_1^2, x_2^2, \dots, x_n^2$, se calculan a continuación los coeficientes $1, c_1, c_2, \dots, c_n$ del polinomio cuyas raíces son los cuadrados de las raíces del polinomio

$$y^n + b_1 y^{n-1} + b_2 y^{n-2} + \dots + b_n,$$

es decir, $x_1^4, x_2^4, \dots, x_n^4$. Luego, y de forma análoga, se obtienen los coeficientes $1, d_1, d_2, \dots, d_n$ del polinomio cuyas raíces son $x_1^8, x_2^8, \dots, x_n^8$; luego, del polinomio cuyas raíces son $x_1^{16}, x_2^{16}, \dots, x_n^{16}$, y así sucesivamente.

Aquí consideraremos sólo la idea fundamental del método de Lobachevski; además, nos limitaremos, por razones de simplicidad, al caso en el que todas las raíces de la ecuación sean reales y distintas en valor absoluto. Sea

$$|x_1| > |x_2| > \dots > |x_n|,$$

es decir, sea x_1 la mayor raíz en valor absoluto, x_2 , la segunda, etcétera. Sea N un número suficientemente grande, y sea

$$X^n + A_1 X^{n-1} + A_2 X^{n-2} + \dots + A_n$$

el polinomio cuyas raíces son iguales a N -ésima potencia de las raíces $x_1, x_2, \dots, x_n$ del polinomio dado, es decir,

$$\begin{aligned} -A_1 &= x_1^N + x_2^N + \dots + x_n^N, \\ A_2 &= x_1^N x_2^N + x_1^N x_3^N + \dots + x_{n-1}^N x_n^N, \\ &\dots\dots\dots \\ \pm A_n &= x_1^N x_2^N \dots x_n^N. \end{aligned}$$

Entonces, para N suficientemente grande, en la sucesión de números $|x_1^N|, |x_2^N|, \dots, |x_n^N|$ cada número es tan pequeño comparado con el precedente que en las expresiones de $A_1, A_2, \dots, A_n$ es lícito retener sólo el primer sumando, despreciando la suma de todos los demás. Se obtienen así las fórmulas aproximadas siguientes:

$$\begin{aligned} x_1^N &\approx -A_1, & x_1^N x_2^N &\approx A_2, \\ x_1^N x_2^N x_3^N &\approx A_3, \dots, & x_1^N x_2^N x_3^N \dots x_n^N &\approx \pm A_n, \end{aligned}$$

o bien, dividiendo cada igualdad aproximada por la anterior y extrayendo las raíces N -ésimas, se obtienen las siguientes fórmulas para x_k :

$$x_1 = \sqrt[N]{-A_1}, x_2 = \sqrt[N]{-\frac{A_2}{A_1}}, x_3 = \sqrt[N]{-\frac{A_3}{A_2}}, \dots, x_n = \sqrt[N]{-\frac{A_n}{A_{n-1}}}.$$

Puede demostrarse que basta con extender el cálculo hasta el polinomio cuyos coeficientes, tomados con los signos $+ - + - \dots$, sean iguales —con el grado de precisión requerido— a los cuadrados de los coeficientes correspondientes del polinomio precedente.

Puede encontrarse una exposición detallada del método de Lobachevski en el conocido libro de A. N. Krylov «Conferencias sobre cálculo aproximado».

Capítulo 5

ECUACIONES DIFERENCIALES ORDINARIAS

§1. *Introducción*

Ejemplos de ecuaciones diferenciales

Las ecuaciones que hemos encontrado hasta ahora respondían en su mayor parte a la necesidad de obtener los valores numéricos de ciertas magnitudes. Cuando, por ejemplo, al buscar los máximos y mínimos de funciones se resolvía una ecuación y se encontraban los puntos para los cuales se anulaba la velocidad de variación de una función, o cuando, en el capítulo 4, se consideraba el problema de encontrar las raíces de un polinomio, se trataba siempre de hallar números concretos. Pero en las aplicaciones de la matemática surgen a menudo problemas de una clase cualitativamente diferente: problemas en los que la incógnita es a su vez una función, una ley que expresa la dependencia de ciertas variables respecto de otras. Por ejemplo, al investigar el proceso del enfriamiento de un cuerpo hay que determinar cómo varía la temperatura en el transcurso del tiempo; para describir el movimiento de un planeta o de una estrella debe determinarse la dependencia de sus coordenadas respecto del tiempo, etcétera.

A menudo es posible plantear una ecuación que permite encontrar las funciones desconocidas pedidas, y estas ecuaciones reciben el nombre de ecuaciones funcionales. Su naturaleza puede ser, en general, muy diversa; de hecho puede decirse que ya conocemos el ejemplo más sencillo y primitivo de ecuación funcional: las funciones implícitas.

Sobre el problema de encontrar funciones desconocidas nos ocuparemos en los capítulos 5, 6 y 7. En este capítulo y en el siguiente consideraremos la clase más importantes de ecuaciones que sirve para determinar tales funciones: las llamadas *ecuaciones diferenciales*; esto es, ecuaciones en que además de la función desconocida aparecen también derivadas de distintos órdenes.

Las siguientes ecuaciones pueden servir de ejemplo:

$$\begin{aligned} \frac{dx}{dt} + P(t)x &= Q(t); \quad \frac{d^2x}{dt^2} + m^2x = A \operatorname{sen} \omega t; \quad \frac{d^2x}{dt^2} = tx, \\ \frac{\delta u}{\delta t} &= \frac{\delta^2 u}{\delta x^2}; \quad \frac{\delta^2 u}{\delta t^2} = \frac{\delta^2 u}{\delta x^2}; \quad \frac{\delta^2 u}{\delta x^2} + \frac{\delta^2 u}{\delta y^2} = 0. \end{aligned} \quad (1)$$

En las tres primeras ecuaciones, la función desconocida se nota por la letra x , y la variable independiente, por la t ; en las tres últimas, la función desconocida se nota por la letra u y depende de dos variables, x y t , o x e y .

La enorme importancia de las ecuaciones diferenciales en la matemática, y especialmente en sus aplicaciones, se debe principalmente al hecho de que la investigación de muchos problemas de física y tecnología puede reducirse a la solución de tales ecuaciones.

Los cálculos que requiere la construcción de maquinaria eléctrica o de dispositivos radiotécnicos, el cálculo de trayectorias de proyectiles, la investigación de la estabilidad de aeronaves en vuelo o del curso de una reacción química, todo ello depende de la solución de ecuaciones diferenciales.

Sucede con frecuencia que las leyes físicas que gobiernan un fenómeno se escriben en forma de ecuaciones diferenciales, por lo que éstas, en sí, constituyen una expresión cuantitativa (numérica) de dichas leyes. El lector verá en los siguientes capítulos cómo las leyes de conservación de la masa y de la energía térmica se expresan en forma de ecuaciones diferenciales. Las leyes de la mecánica descubiertas por Newton permiten estudiar el comportamiento de cualquier sistema mecánico por medio de ecuaciones diferenciales.

Ilustrémoslo con un sencillo ejemplo. Consideremos una partícula material de masa m que se mueve a lo largo del eje Ox , y notemos

por x su coordenada en el instante de tiempo t . La coordenada x variará con el tiempo, y el conocimiento del movimiento de la partícula es equivalente al conocimiento de la dependencia funcional de x respecto del tiempo t . Supongamos que el movimiento lo produce una fuerza F cuyo valor depende de la posición de la partícula (definida por la coordenada x), de la velocidad del movimiento $v = dx/dt$ y del tiempo t , es decir, $F = F(x, dx/dt, t)$. Según las leyes de la mecánica, la acción de la fuerza F sobre la partícula produce necesariamente una aceleración $w = d^2x/dt^2$ tal que el producto de w por la masa m de la partícula es igual a la fuerza; así, en todo instante del movimiento se tiene la ecuación

$$m \frac{d^2x}{dt^2} = F(x, \frac{dx}{dt}, t). \quad (2)$$

Esta es la ecuación diferencial que debe satisfacer la función $x(t)$ que describe el comportamiento de la partícula en movimiento. Es sencillamente una representación de las leyes de la mecánica. Su importancia reside en el hecho de que permite reducir el problema mecánico de determinar el movimiento de una partícula al problema matemático de resolver una ecuación diferencial.

Más adelante, en este mismo capítulo, encontrará el lector otros ejemplos que muestran cómo el estudio de diversos procesos físicos puede reducirse a la investigación de ecuaciones diferenciales.

La teoría de las ecuaciones diferenciales comenzó a desarrollarse a finales del siglo XVII, casi simultáneamente con la aparición del cálculo diferencial e integral. En el momento actual, las ecuaciones diferenciales se han convertido en una herramienta poderosa para la investigación de los fenómenos naturales. En la mecánica, la astronomía, la física y la tecnología han sido causa de enorme progreso. Del estudio de las ecuaciones diferenciales del movimiento de los cuerpos celestes dedujo Newton las leyes del movimiento planetario descubiertas empíricamente por Kepler. En 1846 Leverrier predijo la existencia del planeta Neptuno y determinó su posición en el cielo basándose en el análisis numérico de esas mismas ecuaciones.

Para describir en términos generales los problemas de la teoría de las ecuaciones diferenciales, observemos primero que toda ecuación diferencial tiene, en general, no una, sino un número infinito de soluciones; esto es, existe un conjunto infinito de funciones que la satisfacen. Por ejemplo, la ecuación del movimiento de una partícula debe satisfacerse por cualquier movimiento producido por la fuerza

dada $F(x, dx/dt, t)$, independientemente del punto de partida o de la velocidad inicial. Cada uno de estos distintos movimientos de la partícula irá ligado a una dependencia particular de x respecto del tiempo t . Puesto que bajo una fuerza dada F puede haber infinitos movimientos, la ecuación diferencial (2) tendrá un conjunto infinito de soluciones.

Cualquier ecuación diferencial define, en general, toda una clase de funciones que la satisfacen. El problema básico de la teoría consiste en estudiar las funciones que satisfacen la ecuación diferencial. La teoría de estas ecuaciones debe proporcionar una idea suficientemente amplia de las propiedades de todas las funciones que satisfacen la ecuación, requisito que es particularmente importante al aplicar estas ecuaciones a las ciencias naturales. Además, la teoría debe suministrar el medio de encontrar valores numéricos de una función cuando éstos se necesiten en el transcurso de un cálculo. Más adelante se hablará de cómo pueden obtenerse estos valores numéricos.

Si la función desconocida depende de una única variable, la ecuación diferencial se llama *ecuación diferencial ordinaria*. Si la función desconocida depende de varias variables y la ecuación contiene derivadas respecto a alguna o a todas las variables, la ecuación diferencial se llama *ecuación en derivadas parciales*. Las tres primeras ecuaciones de (1) son ordinarias, y las tres últimas, parciales.

La teoría de las ecuaciones en derivadas parciales tiene muchos rasgos peculiares que la hacen esencialmente diferente de las ecuaciones diferenciales ordinarias. Las ideas básicas implícitas en tales ecuaciones se presentarán en el próximo capítulo; en éste sólo se estudiarán las ecuaciones diferenciales ordinarias.

Consideremos algunos ejemplos.

Ejemplo 1

La ley de desintegración del radio dice que la velocidad de desintegración es proporcional a la cantidad inicial de radio. Supongamos que en cierto instante $t = t_0$ se tienen R_0 gramos de radio. Se desea saber la cantidad de radio presente en cualquier instante posterior t .

Sea $R(t)$ la cantidad de radio existente en el instante t . La velocidad de desintegración viene dada por el valor $-(dR/dt)$. Puesto que este valor es proporcional a R , se tiene

$$-\frac{dR}{dt} = kR, \quad (3)$$

donde k es una constante.

Para resolver el problema, es necesario determinar una función que verifique la ecuación diferencial (3). Para ello observemos que la función inversa de $R(t)$ satisface la ecuación

$$-\frac{dt}{dR} = \frac{1}{kR}, \quad (4)$$

puesto que $dt/dR = 1/(dR/dt)$. Por el cálculo integral se sabe que la ecuación (4) la verifica cualquier función de la forma

$$t = -\frac{1}{k} \ln R + C,$$

donde C es una constante arbitraria. De esta relación determinamos R como función de t . Se tiene

$$R = e^{-kt+kC} = C_1 e^{-kt}. \quad (5)$$

Del conjunto de todas las soluciones (5) de la ecuación (3) hay que seleccionar aquella que en $t = t_0$ tome el valor R_0 . Esta solución se obtiene poniendo $C_1 = R_0 e^{kt_0}$.

Desde el punto de vista matemático, la ecuación (3) es la expresión de una ley muy sencilla para la variación de la función R con el tiempo; dice que la velocidad de decrecimiento $-(dR/dt)$ de la función es proporcional al valor de la función R . Tal ley para la velocidad de variación de una función no sólo rige para los fenómenos de desintegración radiactiva, sino también para otros muchos fenómenos físicos.

Exactamente la misma ley se encuentra en el estudio del enfriamiento de un cuerpo, en donde el ritmo de disminución de la cantidad de calor del cuerpo es proporcional a la diferencia de temperatura entre dicho cuerpo y el medio que lo rodea, y también en otros muchos procesos físicos. Así, el campo de aplicación de la ecuación (3) no se reduce, ni mucho menos, al problema particular de la desintegración radiactiva, problema a partir del cual obtuvimos la ecuación.

Ejemplo 2

Consideremos un punto material de masa m moviéndose sobre el eje horizontal Ox en un medio resistente (por ejemplo, un líquido o un gas) bajo la influencia de la fuerza elástica de dos resortes, que

actúan según la ley de Hooke (Fig. 1); esta ley establece que la fuerza elástica está dirigida hacia la posición de equilibrio y es proporcional a la desviación a partir de dicha posición de equilibrio. Supongamos que esta posición corresponde al punto $x = 0$. Entonces la fuerza elástica es igual a $-bx$, donde $b > 0$.

Supondremos que la resistencia del medio es proporcional a la velocidad del movimiento, es decir, igual a $-a(dx/dt)$, donde $a > 0$

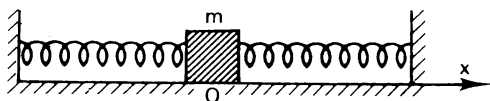


FIG. 1

y el signo «menos» indica que la fuerza resistente actúa en sentido contrario al del movimiento. La experiencia confirma esta hipótesis sobre la resistencia del medio.

A partir de la ley de Newton que dice que el producto de la masa de un punto material por su aceleración es igual a la suma de las fuerzas que actúan sobre él, se tiene

$$m \frac{d^2x}{dt^2} = -bx - a \frac{dx}{dt}. \quad (6)$$

Así, la función $x(t)$, que describe la posición del punto en movimiento en cualquier instante de tiempo t , satisface la ecuación diferencial (6). En una de las últimas secciones estudiaremos las soluciones de esta ecuación.

Si, además de las fuerzas mencionadas, actúa sobre el punto material otra fuerza F , exterior al sistema, entonces la ecuación del movimiento (6) toma la forma

$$m \frac{d^2x}{dt^2} = -bx - a \frac{dx}{dt} + F \quad (6')$$

Ejemplo 3

Un péndulo matemático (o simple) es un punto material de masa m suspendido de un hilo cuya longitud denotaremos por l . Supondremos que el péndulo se mueve siempre sobre el plano del

dibujo (Fig. 2). La fuerza que tiende a llevar al péndulo a la posición vertical OA es el peso mg que actúa sobre el punto material. La posición del péndulo en cualquier instante t viene dada por el ángulo ϕ que forma el hilo con la vertical OA . Tomemos como dirección positiva de ϕ la contraria a las agujas de un reloj. El arco $AA' = l\phi$ es la distancia recorrida por el punto material desde la posición de equilibrio A . La velocidad v del movimiento estará dirigida según la tangente a la circunferencia y tendrá el siguiente valor numérico:

$$v = l \frac{d\phi}{dt}.$$

Para establecer la ecuación del movimiento, descomponemos el peso mg en dos componentes Q y P , la primera en la dirección del radio OA' y la segunda según la tangente al arco AA' . La componente Q no puede afectar al valor numérico de la velocidad v , puesto que está contrarrestada por la resistencia de la suspensión OA' . Sólo la componente P puede afectar al valor de la velocidad v . Esta componente actúa siempre hacia la posición de equilibrio A , es decir, hacia un decrecimiento de ϕ si el ángulo es positivo, y hacia un crecimiento de ϕ cuando es negativo. El valor de P es $-mg \sin \phi$, por lo que la ecuación del movimiento del péndulo es

$$m \frac{dv}{dt} = -mg \sin \phi$$

o bien

$$\frac{d^2\phi}{dt^2} = -\frac{g}{l} \sin \phi. \quad (7)$$

Es interesante observar que las soluciones de esta ecuación no pueden expresarse mediante una combinación finita de funciones elementales.

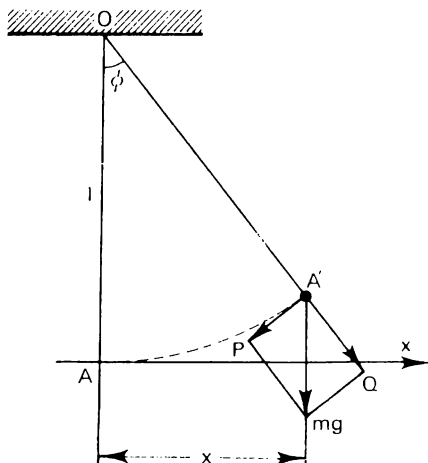


FIG. 2

El conjunto de las funciones elementales es demasiado pequeño para dar una descripción exacta ni siquiera de procesos físicos tan sencillos como el de la oscilación de un péndulo simple. Más tarde veremos que las ecuaciones diferenciales resolubles mediante funciones elementales no son muy numerosas, y por ello ocurre frecuentemente que el estudio de las ecuaciones diferenciales que aparecen en la física o en la mecánica lleva a introducir nuevas clases de funciones,

que se convierten en objeto de estudio, ampliando así el arsenal de funciones que pueden emplearse en la resolución de problemas prácticos.

Limitémonos ahora a oscilaciones pequeñas del péndulo, para las cuales podemos suponer (cometiendo un error muy pequeño) que el arco AA' es igual a su proyección x sobre el eje horizontal Ox , y que $\sin \phi$ es igual a ϕ . Entonces $\phi \approx \sin \phi = x/l$, y la ecuación de movimiento del péndulo tomará la forma más sencilla

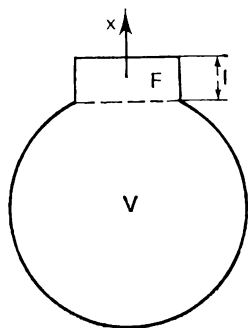


FIG. 3

$$\frac{d^2x}{dt^2} = -\frac{g}{l}x. \quad (8)$$

Más adelante veremos que esta ecuación es resoluble por medio de funciones trigonométricas y que utilizándolas podemos llegar a describir con suficiente exactitud las «pequeñas oscilaciones» del péndulo.

Ejemplo 4

El resonador acústico de Helmholtz (Fig. 3) consiste en un recipiente V lleno de aire, cuyo volumen es v , con un cuello cilíndrico F . Aproximadamente, se puede considerar el aire del cuello del recipiente como un tapón de masa

$$m = \rho sl, \quad (9)$$

donde ρ es la densidad del aire; s , el área de la sección transversal del cuello, y l , su longitud. Si suponemos que esta masa de aire se desplaza una longitud x a partir de su posición de equilibrio, entonces la presión del aire en el recipiente de volumen v incrementará su valor inicial p en una cantidad que llamaremos Δp .

Supondremos que la presión p y el volumen v siguen la ley adia-

bática $pv^k = C$. Entonces, despreciando magnitudes de orden superior, tendremos

$$\Delta p \cdot v^k + pk v^{k-1} \cdot \Delta v = 0$$

e

$$\Delta p = -kp \frac{\Delta v}{v} = -\frac{kps}{v} x. \quad (10)$$

(En este caso, $\Delta v = sx$). La ecuación del movimiento de la masa de aire del cuello puede escribirse como:

$$m \frac{d^2 x}{dt^2} = \Delta p \cdot s. \quad (11)$$

$\Delta p \cdot s$ es la fuerza que el gas del recipiente ejerce sobre la columna de aire del cuello. De (10) y (11) se tiene

$$\rho l \frac{d^2 x}{dt^2} = -\frac{kps}{v} x, \quad (12)$$

donde ρ, p, v, l, k y s son constantes.

Ejemplo 5

También aparece una ecuación de la forma (6) al estudiar las oscilaciones eléctricas en un circuito oscilante. El diagrama del circuito está representado en la figura 4. Dicho circuito consta de un condensador de capacidad C en serie con una bobina de inductancia L y con una resistencia R . En un instante determinado el condensador tendrá un cierto voltaje entre sus terminales. En ausencia de inductancia en el circuito, se establecerá una corriente hasta que los dos terminales del condensador estén al mismo potencial. La presencia de una inductancia altera la situación, puesto

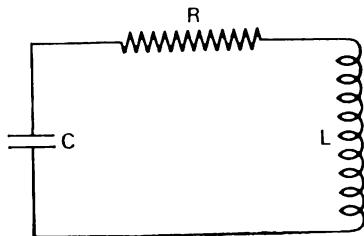


FIG. 4

que el circuito generará ahora oscilaciones eléctricas. Para encontrar una fórmula para estas oscilaciones, denotaremos por $v(t)$, o sencillamente por v , el voltaje o diferencia de potencial del condensador en el instante t ; por $I(t)$, la corriente en el mismo instante, y por R , la

resistencia. Según leyes conocidas de la física, $I(t)R$ permanece siempre igual a la fuerza electromotriz total, que es la suma del voltaje a través del condensador y la inductancia $-L(dI/dt)$. Así,

$$IR = -v - L \frac{dI}{dt}. \quad (13)$$

Notemos por $Q(t)$ la carga del condensador en el instante t . Entonces la corriente en el circuito será, en cada instante, igual a dQ/dt . La diferencia de potencial $v(t)$ a través del condensador es igual a $Q(t)/C$. Así, $I = dQ/dt = C(dv/dt)$, y la ecuación (13) puede transformarse en

$$LC \frac{d^2v}{dt^2} + RC \frac{dv}{dt} + v = 0. \quad (14)$$

Ejemplo 6

La figura 5 muestra el diagrama del circuito de un tubo electrónico generador de oscilaciones electromagnéticas. El circuito oscilante, que consta de una capacidad C en paralelo con una resistencia R y una inductancia L , representa el sistema oscilador básico. La bobina L' y el tubo dibujado en el centro de la figura 5 forman el sistema de «realimentación» y conectan el generador de energía (la batería B) al circuito L - R - C ; K es el cátodo del tubo; A , la placa, y S , la rejilla. En un circuito L - R - C de este tipo aparecerán «auto-oscilaciones». En todo sistema real en estado oscilante, parte de la energía se transforma en calor o se disipa de alguna otra forma a los cuerpos que lo rodean; por ello, para mantener un estado estacionario de oscilación es necesario disponer de un generador exterior de energía. Las auto-oscilaciones difieren de otros procesos oscilatorios en que, para mantener un estado oscilatorio estacionario del sistema, no es preciso que el generador exterior sea periódico. Los sistemas auto-oscilantes se construyen de tal modo que un generador constante de energía, en este caso la batería B , mantenga un estado oscilatorio estacionario. Ejemplos de sistemas auto-oscilantes son un reloj, un timbre eléctrico, la cuerda y el arco de un instrumento musical, la voz humana, etc.

La corriente $I(t)$ en el circuito oscilante L - R - C satisface la ecuación

$$L \frac{dI}{dt} + RI + v = M \frac{dI_a}{dt}, \quad (15)$$

donde $v = v(t)$ es el voltaje del condensador en el instante t ; $I_a(t)$ es la corriente de placa a través de la bobina L' ; M es el coeficiente de acoplamiento entre las bobinas L y L' . Comparándola con la ecuación (13), la (15) contiene el término adicional $M(dI_a/dt)$.

Supondremos que la corriente de placa $I_a(t)$ depende sólo del voltaje entre la rejilla S y el cátodo del tubo (es decir, despreciaremos la reactancia del ánodo), de modo que este voltaje es igual al vol-

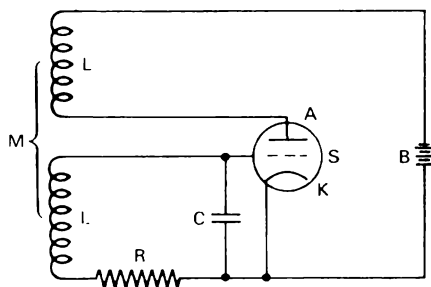


FIG. 5

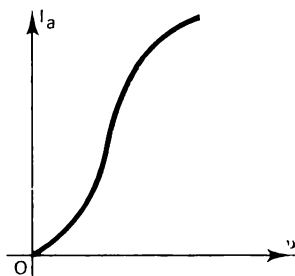


FIG. 6

taje $v(t)$ del condensador C . El carácter de la dependencia funcional de I_a respecto de v se observa en la figura 6. La curva se dibuja generalmente como una parábola cúbica, y una expresión aproximada de ella será:

$$I_a = a_1 v + a_2 v^2 + a_3 v^3.$$

Sustituyendo esta expresión en el segundo miembro de la ecuación (15), y teniendo en cuenta que

$$\frac{dv}{dt} = I,$$

se obtiene para v la ecuación siguiente:

$$L \frac{d^2 v}{dt^2} + [R - M(a_1 + 2a_2 v + 3a_3 v^2)] \frac{dv}{dt} + v = 0. \quad (16)$$

En los ejemplos considerados, la búsqueda de ciertas expresiones características de un proceso físico dado se reduce a la búsqueda de soluciones de ecuaciones diferenciales ordinarias.

Problemas en la teoría de las ecuaciones diferenciales

A continuación damos definiciones precisas. *Una ecuación diferencial ordinaria de orden n en una función incógnita y es una relación de la forma*

$$F[x, y(x), y'(x), y''(x), \dots, y^{(n)}(x)] = 0 \quad (17)$$

entre la variable independiente x y las expresiones

$$y(x), y'(x) = \frac{dy}{dx}, y''(x) = \frac{d^2y}{dx^2}, \dots, y^{(n)}(x) = \frac{d^ny}{dx^n}.$$

El orden de una ecuación diferencial es el orden de la máxima derivada de la función que aparezca en la ecuación diferencial. Así, la ecuación del ejemplo 1 es de primer orden, y las de los ejemplos 2, 3, 4, 5 y 6 son de segundo orden.

Una función $\phi(x)$ se dice *solución de la ecuación diferencial* (17) si, al sustituir y por $\phi(x)$, y' por $\phi'(x)$, $\dots$, $y^{(n)}$ por $\phi^{(n)}(x)$, resulta una identidad.

Los problemas de la física y la tecnología conducen a menudo a un sistema de ecuaciones diferenciales ordinarias con varias funciones incógnitas, dependiendo todas de la misma variable y de sus derivadas respecto a esa variable.

Para concretar más, las explicaciones que siguen se referirán principalmente a una ecuación diferencial ordinaria de segundo orden como máximo. Con este ejemplo se pueden explicar las propiedades esenciales de todas las ecuaciones diferenciales ordinarias y de los sistemas de ecuaciones en los cuales el número de funciones desconocidas es igual al de ecuaciones.

Hemos hablado anteriormente del hecho de que, por regla general, toda ecuación diferencial no tiene una, sino un conjunto infinito de soluciones. Ilustraremos esto en primer lugar mediante consideraciones intuitivas basadas en los ejemplos definidos por las ecuaciones (2-6). En todos ellos, la correspondiente ecuación diferencial está ya completamente definida por el planteamiento físico del sistema. Pero en cada uno de estos sistemas puede haber muchos movimientos distintos. Por ejemplo, está perfectamente claro que un péndulo, descrito por la ecuación (8), puede oscilar con muchas amplitudes distintas. Cada una de estas distintas oscilaciones del péndulo irá ligada a una solución diferente de la ecuación (8), por lo que deben

existir un número infinito de soluciones. Puede demostrarse que la ecuación (8) la satisface cualquier función de la forma

$$x = C_1 \cos \sqrt{\frac{g}{l}} t + C_2 \sin \sqrt{\frac{g}{l}} t, \quad (18)$$

donde C_1 y C_2 son constantes arbitrarias.

Físicamente también está claro que el movimiento del péndulo sólo quedará completamente determinado si se especifica, para algún instante t_0 , el valor (inicial) x_0 de x (desplazamiento inicial del punto material a partir de su posición de equilibrio) y la velocidad inicial del movimiento $x'_0 = (dx/dt) |_{t=0}$. Estas condiciones iniciales determinan las constantes C_1 y C_2 en la fórmula (18).

Exactamente del mismo modo, las ecuaciones diferenciales que hemos encontrado en otros ejemplos tendrán infinitas soluciones.

En general puede demostrarse, bajo hipótesis muy amplias, que la ecuación diferencial (17) de orden n tiene infinitas soluciones. Con más precisión: si para un «valor inicial» de la variable independiente asignamos un «valor inicial» a la función desconocida y a todas sus derivadas hasta la de orden $n - 1$, existe entonces una solución de la ecuación (17) con estos valores iniciales asignados. Puede demostrarse también que tales condiciones iniciales determinan completamente la solución, por lo que sólo existe *una* solución que satisfaga las condiciones iniciales. Esta cuestión se discutirá más tarde con detalle. Ahora es esencial advertir que los valores iniciales de la función y de las $n - 1$ primeras derivadas pueden fijarse arbitrariamente. Cada elección de estos n valores definirá un «estado inicial» para la solución buscada.

Para hallar una fórmula que incluya, a ser posible, todas las soluciones de una ecuación diferencial de orden n , aquélla deberá contener n constantes independientes, lo que permitirá imponer n condiciones iniciales. Las soluciones de una ecuación diferencial de orden n que contienen n constantes arbitrarias independientes recibe el nombre de *solución general* de la ecuación. Por ejemplo, una solución general de (8) viene dada por la fórmula (18), con dos constantes arbitrarias; una solución general de la ecuación (3) viene dada por la fórmula (5).

Trataremos ahora de formular en un esquema muy general los problemas a los que se aplica la teoría de las ecuaciones diferenciales. Son muchos y muy variados, y sólo indicaremos los más importantes.

Si la ecuación diferencial viene acompañada de sus condiciones iniciales, su solución está completamente determinada. La obtención de fórmulas que permitan calcular la solución en forma explícita será uno de los primeros problemas de la teoría. Tales fórmulas sólo podrán obtenerse en casos muy sencillos, pero, si se encuentran, serán de gran ayuda en el cálculo y estudio de la solución.

La teoría deberá suministrar algún medio para obtener propiedades referentes al comportamiento de una solución: si es monótona u oscilatoria, si es periódica o se aproxima a una función periódica, etc.

Si variamos los valores iniciales de la función incógnita y de sus derivadas, esto es, si modificamos el estado inicial del sistema físico, entonces variará también la solución, puesto que el proceso físico tendrá lugar ahora de modo diferente. La teoría deberá suministrar la posibilidad de predecir cuál será este cambio. En particular, para pequeños cambios en los valores iniciales, ¿se modificará también la solución en una cantidad pequeña, siendo por tanto estable a este respecto, o podrá ocurrir que pequeños cambios en las condiciones iniciales den lugar a grandes modificaciones en la solución, de modo que ésta última sea inestable?

También deberá proporcionar un cuadro cualitativo, y si fuese posible cuantitativo, del comportamiento, no sólo de las soluciones particulares de una ecuación, sino también de la solución general.

En la construcción de maquinaria se plantea a menudo la cuestión de elegir los parámetros que caracterizan a un aparato o máquina, de modo tal que garanticen un funcionamiento satisfactorio. Los parámetros de un aparato aparecen en forma de ciertas magnitudes en la ecuación diferencial correspondiente. La teoría deberá ayudar a aclarar qué le sucederá a la solución de la ecuación (al aparato) si se modifica la ecuación diferencial (los parámetros del aparato).

Finalmente, cuando sea necesario realizar un cálculo se necesitará encontrar la solución numérica de una ecuación, y la teoría deberá proporcionar al ingeniero y al físico los métodos más rápidos y económicos de calcular las soluciones.

§2. Ecuaciones diferenciales lineales con coeficientes constantes

Para ciertas clases importantes de ecuaciones diferenciales ordinarias la solución general se puede expresar en términos de funciones sencillas y bien conocidas. Una de estas clases es la de aquellas ecua-

ciones diferenciales con coeficientes constantes que son lineales respecto a la función desconocida y a sus derivadas (en pocas palabras, lineales). Las ecuaciones diferenciales (3), (6), (8) y (14) son ejemplos de tales ecuaciones. Una ecuación lineal se llama homogénea si carece de términos que contengan la variable desconocida, y no homogénea, si existe alguno de dichos términos.

Ecuaciones lineales homogéneas de segundo orden con coeficientes constantes

Son ecuaciones de la forma

$$m \frac{d^2x}{dt^2} + a \frac{dx}{dt} + bx = 0, \quad (6)$$

donde m , a y b son constantes. Supondremos que m es positivo; esto no restringirá la generalidad, puesto que siempre se puede lograr que sea positivo cambiando, si es necesario, el signo de todos los coeficientes, siempre que $m \neq 0$, lo cual supondremos que se cumple.

Busquemos una solución de esta ecuación en forma de una función exponencial $e^{\lambda t}$, y elijamos la constante λ de modo que la función $x = e^{\lambda t}$ satisfaga la ecuación. Sustituyendo $x = e^{\lambda t}$, $dx/dt = \lambda e^{\lambda t}$ y $d^2x/dt^2 = \lambda^2 e^{\lambda t}$ en el segundo miembro de la ecuación (6), se obtiene

$$e^{\lambda t}(m\lambda^2 + a\lambda + b).$$

Así, para que $x(t) = e^{\lambda t}$ sea solución de la ecuación (6) es necesario y suficiente que

$$m^2\lambda^2 + a\lambda + b = 0. \quad (19)$$

Si λ_1 y λ_2 son dos raíces reales de la ecuación (19), entonces se demuestra fácilmente que una solución de la ecuación (6) viene dada por toda función de la forma

$$x = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}, \quad (20)$$

donde C_1 y C_2 son constantes arbitrarias.

Más adelante se demostrará que la fórmula (20) da todas las soluciones de la ecuación (6) en el caso de que la ecuación (19) tenga raíces reales distintas.

Se advierten las siguientes propiedades importantes de las soluciones de la ecuación (6):

1. La suma de dos soluciones es también una solución.

2. Una solución multiplicada por una constante es también solución.

En el caso de que λ_1 sea una raíz múltiple de la ecuación (19), es decir, $m\lambda_1^2 + a\lambda_1 + b = 0$ y $2m\lambda_1 + a = 0$ ¹, una solución de la ecuación (6) vendrá dada también por la función $te^{\lambda_1 t}$, puesto que, si sustituimos esta función y sus derivadas en el segundo miembro de la ecuación (6), se obtiene

$$te^{\lambda_1 t}(m\lambda_1^2 + a\lambda_1 + b) + e^{\lambda_1 t}(2m\lambda_1 + a),$$

expresión que es idénticamente nula, como se deduce de las ecuaciones anteriores.

La solución general de la ecuación (6) en este caso es de la forma

$$x = C_1 e^{\lambda_1 t} + C_2 t e^{\lambda_1 t}. \quad (21)$$

Supongamos ahora que la ecuación (19) tiene raíces complejas. Estas raíces serán complejas conjugadas una de otra, puesto que m , a y b son números reales. Sea $\lambda = a \pm i\beta$. La ecuación

$$m(a + i\beta)^2 + a(a + i\beta) + b = 0$$

es equivalente a las dos ecuaciones

$$ma^2 - m\beta^2 + aa + b = 0 \quad \text{y} \quad 2ma\beta + a\beta = 0. \quad (22)$$

Es fácil demostrar que en este caso las funciones $x = e^{at} \cos \beta t$ y $x = e^{at} \sin \beta t$ son soluciones de la ecuación (6). Así, por ejemplo, sustituyendo la función $x(t) = e^{at} \cos \beta t$ y sus derivadas en el primer miembro de la ecuación (6), se obtiene

$$e^{at} \cos \beta t (ma^2 - m\beta^2 + aa + b) - e^{at} \sin \beta t (2ma\beta + a\beta).$$

De la ecuación (22) se deduce que esta expresión es idénticamente nula.

La solución general de la ecuación (6), cuando la (19) tiene raíces

¹ La suma de las raíces λ_1 y λ_2 de la ecuación de segundo grado (19) es $\lambda_1 + \lambda_2 = -a/m$, y si las raíces son iguales, es decir, si $\lambda_1 = \lambda_2$, entonces es válida la segunda de las ecuaciones anteriores.

complejas, es de la forma

$$x = C_1 e^{at} \operatorname{sen} \beta t + C_2 e^{at} \cos \beta t, \quad (23)$$

donde C_1 y C_2 son constantes arbitrarias.

De este modo, si se conocen las raíces de la ecuación (19), llamada *ecuación característica*, se puede escribir la solución general de la ecuación (6).

Observemos que la solución general de una ecuación homogénea lineal de orden n con coeficientes constantes

$$a_n \frac{d^n x}{dt^n} + a_{n-1} \frac{d^{n-1} x}{dt^{n-1}} + \cdots + a_1 \frac{dx}{dt} + a_0 x = 0$$

puede escribirse de modo similar en forma de un polinomio en funciones trigonométricas y exponenciales, siempre que se conozcan las raíces de la ecuación algebraica

$$a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_0 = 0,$$

que también se llama *ecuación característica*. Así, el problema de integrar una ecuación diferencial ordinaria lineal con coeficientes constantes se reduce a un problema algebraico.

Demostraremos ahora que las fórmulas (20), (21) y (23) dan todas las soluciones de la ecuación (6). Observemos que C_1 y C_2 pueden elegirse siempre de modo que la función $x(t)$ satisfaga condiciones iniciales arbitrarias $x(t_0) = x_0$, $x'(t_0) = x'_0$. Para ello hay que determinar C_1 y C_2 en el sistema de ecuaciones

$$\begin{aligned} x_0 &= C_1 e^{\lambda_1 t_0} + C_2 e^{\lambda_2 t_0} \\ x'_0 &= \lambda_1 C_1 e^{\lambda_1 t_0} + \lambda_2 C_2 e^{\lambda_2 t_0} \end{aligned}$$

en el caso de la fórmula (20), o por dos ecuaciones semejantes en el caso de las fórmulas (21) y (23). Evidentemente, de existir una solución de la ecuación (6) que no estuviera comprendida entre las soluciones que se han obtenido, existirían entonces dos soluciones distintas de la ecuación (6) que satisfarían las mismas condiciones iniciales. Su diferencia $x_1(t)$ no sería idénticamente nula y satisfaría las condiciones iniciales nulas $x_1(t_0) = 0$, $x'_1(t_0) = 0$. Demostraremos que una solución de la ecuación (6) que satisfaga las condiciones iniciales nulas sólo puede ser $x_1(t) = 0$. Demostraremos primero esto

suponiendo que $m > 0$, $a > 0$, y $b > 0$. Para ello multipliquemos los dos miembros de la ecuación

$$m \frac{d^2 x_1}{dt^2} + a \frac{dx_1}{dt} + bx_1 = 0 \quad (24)$$

por $2(dx_1/dt)$. Puesto que

$$2 \frac{dx_1}{dt} \cdot \frac{d^2 x_1}{dt^2} = \frac{d}{dt} \left(\frac{dx_1}{dt} \right)^2 \quad \text{y} \quad 2x_1(t) \frac{dx_1}{dt} = \frac{d}{dt} (x_1^2),$$

la ecuación (24) puede ponerse en la forma

$$\frac{d}{dt} \left[m \left(\frac{dx_1}{dt} \right)^2 \right] + 2a \left(\frac{dx_1}{dt} \right)^2 + b \frac{d}{dt} (x_1^2) = 0.$$

Integrando esta identidad entre t_0 y t , se obtiene

$$m \left(\frac{dx_1}{dt} \right)^2 + 2a \int_{t_0}^t \left(\frac{dx_1}{dt} \right)^2 dt + bx_1^2(t) = 0.$$

Esta ecuación sólo tiene sentido si $x_1(t) \equiv 0$. Pues, de otro modo, para $t = t_0$, evidentemente se tendría en el primer miembro una cantidad positiva y cero en el segundo, con una situación semejante para $t < t_0$.

Para establecer la proposición para cualesquiera coeficientes constantes m , a y b , consideramos la función $y_1(t) = x_1(t) e^{-at}$ que, como es fácil demostrar, satisface también las condiciones de contorno nulas. Si se elige un valor de $a > 0$ suficientemente grande, la función $y_1(t)$ satisfará una ecuación de la forma (6) para $a > 0$, $b > 0$ y $m > 0$. Esta ecuación se obtiene fácilmente sustituyendo la función $x_1(t) = y_1(t) e^{at}$ y sus derivadas en la ecuación (6). Entonces, como se demostró anteriormente, se tiene $y_1(t) \equiv 0$, lo que significa $x_1(t) = y_1(t) e^{at} \equiv 0$.

Así, pues, hemos demostrado que las fórmulas (20), (21) y (23) proporcionan todas las soluciones de la ecuación (6).

Veamos qué información proporcionan estas fórmulas sobre el carácter de las soluciones de la ecuación (6). Para ello analicemos la expresión

$$\lambda_{1,2} = -\frac{a}{2m} \pm \sqrt{\frac{a^2}{4m^2} - \frac{b}{m}} \quad (25)$$

que da las raíces de la ecuación (19). De acuerdo con los ejemplos de aplicación que condujeron a la ecuación (6), supondremos que $m > 0$, $a \geq 0$ y $b > 0$.

Caso 1

$a^2 > 4bm$. Las dos raíces de la ecuación característica (19) son reales, negativas y distintas. En este caso la función $x(t)$ dada por la fórmula (20) es una solución general de la ecuación (6). Todas las funciones definidas por esta fórmula junto con sus primeras derivadas tienden a cero para $t \rightarrow +\infty$, y sólo hay un valor de t , como máximo, para el que se anulan. Por tanto, la función $x(t)$ no tiene más de un máximo o mínimo. Físicamente, esto significa que la resistencia del medio es suficientemente grande para impedir las oscilaciones. El punto en movimiento no puede pasar por la posición de equilibrio $x = 0$ más de una vez. A partir de ese momento, después de alcanzar una cierta distancia máxima del punto $x = 0$, empieza a aproximarse lentamente a este punto aunque sin volver a pasar por él.

Caso 2

$a^2 = 4bm$. Las dos raíces de la ecuación (19) son iguales y la solución general de la ecuación (6) viene dada por la fórmula (21). En este caso también todas las soluciones $x(t)$ y sus primeras derivadas tienden a cero para $t \rightarrow +\infty$. $x(t)$ y $x'(t)$ no pueden anularse más de una vez. El carácter del movimiento del punto material de abscisa $x(t)$ es el mismo que en el primer caso.

Caso 3

$a^2 < 4bm$. Las raíces de la ecuación característica (19) tienen parte imaginaria no nula. La solución general de la ecuación (6) viene dada por la fórmula (23). El punto x realiza oscilaciones a lo largo del eje x con un período constante $2\pi/\beta$, que es el mismo para todas las soluciones de (6), y con una amplitud Ce^{at} , donde $\alpha = -(a/2m)$.

Aquellas oscilaciones de un sistema físico que tienen lugar sin la acción de una fuerza exterior se llaman oscilaciones características (auto-oscilaciones u oscilaciones propias) del sistema. De la discusión anterior se sigue que el período de tales oscilaciones para los sistemas estudiados en los ejemplos 2, 3, 4 y 5, sólo depende de la estructura del sistema y será el mismo para todas las oscilaciones que puedan aparecer en él. Dicho período es $2\pi\sqrt{b/m} - a^2/4m^2$ en el

ejemplo 2; $2\pi\sqrt{kps/vql}$ en el ejemplo 4, y $2\pi\sqrt{1/LC - R^2/4L^2}$ en el ejemplo 5.

Si $a = 0$, es decir, si el medio no ofrece resistencia al movimiento, la amplitud de las oscilaciones es constante; el punto oscila armónicamente. Pero si $a > 0$, es decir, si el medio ofrece resistencia al movimiento, aunque esta resistencia sea pequeña ($a^2 < 4bm$), la amplitud de las oscilaciones tiende a cero y las oscilaciones terminan por desaparecer.

Finalmente, la solución $x(t) \equiv 0$ de la ecuación (6) indica en todos los casos un estado de reposo del punto x en la posición $x = 0$, que se llama posición de equilibrio.

Si las partes reales de ambas raíces de la ecuación (19) son negativas, entonces, a partir de las fórmulas (20), (21), (23), se deduce que todas las soluciones de la ecuación (6) y sus derivadas tienden a cero para $t \rightarrow +\infty$; es decir, las oscilaciones desaparecen en el transcurso del tiempo.

Sin embargo, con que la parte real de una de las raíces de la ecuación (19) sea positiva, hay soluciones de la ecuación (6) que no tienden a cero para $t \rightarrow +\infty$, por lo que algunas de las soluciones de la ecuación (6) ni siquiera estarían acotadas para $t \rightarrow +\infty$. Tal caso sólo es posible con $m > 0$ si b o a es negativo. Físicamente, esto sólo puede ocurrir en el caso en el que la fuerza elástica no atraiga al punto x hacia la posición de equilibrio, sino que lo repela, o cuando la resistencia del medio es negativa. Tales casos no pueden darse en los ejemplos físicos considerados al comienzo de este capítulo, pero son perfectamente posibles en otros modelos físicos.

Si la parte real de las raíces λ_1 y λ_2 de la ecuación (19) es igual a cero —lo que sólo puede ocurrir si el coeficiente a de la ecuación (19) es cero—, entonces para $a = 0$ el punto $x(t)$, como puede verse en la fórmula (23), realiza oscilaciones armónicas de amplitud y velocidad acotadas.

Ecuaciones lineales no homogéneas con coeficientes constantes

Consideremos en detalle la ecuación

$$m \frac{d^2x}{dt^2} + a \frac{dx}{dt} + bx = A \cos \omega t. \quad (26)$$

Esta es la ecuación de las oscilaciones lineales de un punto material bajo la acción de una fuerza elástica, de la resistencia de un medio y de fuerza periódica externa $A \cos \omega t$ (ver ecuación (6') en §1).

La ecuación (26) es una ecuación lineal no homogénea, y (6) es la correspondiente ecuación homogénea.

Busquemos la solución general de la ecuación (26).

Para ello observemos que la suma de una solución de la ecuación no homogénea y de una solución de la ecuación homogénea correspondiente es también solución de la ecuación lineal no homogénea. Así, para encontrar una solución general de la ecuación (26) es suficiente encontrar una solución particular cualquiera. La solución general de la ecuación (26) vendrá entonces dada como suma de esta solución particular y de la solución general de la correspondiente ecuación homogénea.

Es natural esperar que el movimiento siga el ritmo de la fuerza periódica externa, por lo cual buscaremos una solución particular de la ecuación (26) de la forma $x = B \cos(\omega t + \delta)$, donde B y δ son constantes a determinar. Intentaremos determinar B y δ de modo que la función $x = B \cos(\omega t + \delta)$ satisfaga la ecuación (26). Calculando las derivadas $dx/dt = -B\omega \sin(\omega t + \delta)$ y $d^2x/dt^2 = -B\omega^2 \cos(\omega t + \delta)$ y sustituyéndolas en la ecuación (26), obtenemos

$$m[-B\omega^2 \cos(\omega t + \delta)] + a[-B\omega \sin(\omega t + \delta)] + bB \cos(\omega t + \delta) = A \cos \omega t.$$

Aplicando fórmulas conocidas, tenemos

$$B[(b - m\omega^2) \cos(\omega t + \delta) - a\omega \sin(\omega t + \delta)] = \\ = B \sqrt{(b - m\omega^2)^2 + a^2\omega^2} \cos(\omega t + \delta') = A \cos \omega t,$$

donde $\delta' = \delta + \gamma$ y $\gamma = \arctg a\omega/(b - m\omega^2)$. Evidentemente, si ponemos

$$\delta = -\arctg \frac{a\omega}{b - m\omega^2} \quad \text{y} \quad B = \frac{A}{\sqrt{(b - m\omega^2)^2 + a^2\omega^2}},$$

la función $x = B \cos(\omega t + \delta)$ satisfará la ecuación (26).

Una solución de la forma $B \cos(\omega t + \delta)$ existirá siempre que $(b - m\omega^2)^2 + a^2\omega^2 \neq 0$. En el caso $(b - m\omega^2)^2 + a^2\omega^2 = 0$, es decir, cuando $a = 0$ y $b = m\omega^2$, la ecuación (26) toma la forma

$$m \frac{d^2x}{dt^2} + m\omega^2 x = A \cos \omega t.$$

Una solución particular en este caso, como se deduce fácilmente, es $x = (At/2\sqrt{mb}) \operatorname{sen} \omega t$.

Las soluciones de la ecuación no homogénea (26) se llaman oscilaciones forzadas. El factor $\phi(\omega) = 1/\sqrt{(b - m\omega^2)^2 + a^2\omega^2}$ caracteriza la relación entre la amplitud B de la oscilación forzada y la amplitud A de la fuerza exterior. La gráfica de la función $\phi(\omega)$ se llama curva de resonancia. La frecuencia ω para la cual $\phi(\omega)$ alcanza su máximo se llama frecuencia de resonancia. Calculémosla. Si $\phi(\omega)$ tiene su máximo en $\omega_1 \neq 0$, entonces para este valor de ω la derivada $\phi'(\omega)$ se anula, es decir, $-4(b - m\omega_1^2)m\omega_1 + 2a^2\omega_1 = 0$, por lo que $\omega_1 = \sqrt{b/m - a^2/2m^2}$. Para este valor de ω_1 ,

$$\phi(\omega_1) = \frac{1}{a\sqrt{b/m - a^2/4m^2}}.$$

Se deduce, por tanto, que la amplitud de las oscilaciones forzadas para $\omega = \omega_1$ es mayor para valores pequeños de a . Para a muy pequeño, la frecuencia ω_1 está muy próxima al valor $\sqrt{b/m}$, es decir, a la frecuencia de las oscilaciones libres. Para $a = 0$ y $b = m\omega^2$, como vimos, la oscilación forzada es de la forma

$$x = \frac{At}{2\sqrt{mb}} \operatorname{sen} \omega t,$$

es decir, la amplitud de esta oscilación crece sin límite cuando $t \rightarrow +\infty$, situación que representa el significado matemático de la resonancia. La resonancia aparecerá si el período de la fuerza exterior es el mismo que el de una de las oscilaciones características del sistema. En el mundo real, cuando el período de la fuerza exterior se aproxima mucho al de las oscilaciones características, las oscilaciones del sistema pueden hacerse extremadamente grandes.

La posibilidad de grandes oscilaciones se utiliza a menudo en radiotécnica para distintas clases de amplificadores. Pero las grandes oscilaciones pueden también conducir a la rotura de estructuras tales como puentes o máquinas. Por ello es muy importante tener en cuenta la posibilidad de resonancia o de oscilaciones próximas a ella.

De las observaciones anteriores se deduce que cualquier solución de la ecuación (26) puede escribirse como suma de la oscilación forzada que acabamos de encontrar y de una de las soluciones de la

ecuación homogénea dada por las fórmulas (20), (21) y (23). Para $a > 0$ y $b > 0$ la solución de la ecuación homogénea tiende a cero para $t \rightarrow +\infty$, es decir, cualquier movimiento termina aproximándose a las oscilaciones forzadas. Si $a = 0$ y $b > 0$, la oscilación forzada se superpone a una oscilación característica no amortiguada del sistema. Para $b = m\omega^2$ y $a = 0$, se tiene la resonancia.

Cuando actúa una fuerza periódica externa $f(t)$ sobre el sistema, las oscilaciones forzadas de éste pueden obtenerse del modo siguiente: $f(t)$ se puede representar con una aproximación suficiente como parte de una serie trigonométrica ²

$$\sum_{i=1}^n (a_i \cos \omega_i t + b_i \sin \omega_i t). \quad (27)$$

Una vez halladas las oscilaciones forzadas correspondientes a cada término de esta suma, las oscilaciones correspondientes a la fuerza $f(t)$ se encontrarán sumando las oscilaciones correspondientes a los distintos términos de la suma (27). Si cualquiera de estas frecuencias es idéntica a la de la oscilación característica del sistema, habrá resonancia.

§3. *Observaciones generales sobre la formación y solución de las ecuaciones diferenciales*

Son pocas las ecuaciones diferenciales cuyas soluciones puedan expresarse todas ellas explícitamente en términos de funciones sencillas, como es el caso de las ecuaciones lineales con coeficientes constantes. Se pueden dar ejemplos sencillos de ecuaciones diferenciales cuya solución general no se puede expresar por medio de un número finito de integrales de funciones conocidas, o, como también se dice, de cuadraturas.

Como demostró Liouville en 1841, la solución de la ecuación de Riccati de la forma $dy/dx + ay^2 = x^2$, para $a > 0$, no puede expresarse como combinación finita de integrales de las funciones elementales. Por ello es importante desarrollar métodos de aproximación para las soluciones de las ecuaciones diferenciales que se puedan aplicar a una amplia gama de ecuaciones.

² Ver capítulo 12, §7.

El hecho de que en tales casos no encontremos soluciones exactas sino sólo aproximadas no debe preocuparnos. En primer lugar, estas soluciones pueden calcularse, al menos en principio, con cualquier aproximación deseada. En segundo lugar, hay que tener en cuenta que en la mayoría de los casos las ecuaciones diferenciales que describen un proceso físico no son totalmente exactas, como puede verse en cada uno de los ejemplos discutidos en el §1.

Un ejemplo particularmente interesante lo proporciona la ecuación (12) para el resonador acústico. Para deducir esta ecuación se despreció la compresibilidad del aire en el cuello del recipiente, así como el movimiento del aire en el mismo. De hecho, el movimiento del aire en el cuello pone en movimiento la masa de aire del recipiente, pero estos dos movimientos tienen velocidades y desplazamientos diferentes. En el cuello, el desplazamiento de las partículas de aire es considerablemente mayor que en el recipiente. De este modo despreciamos el movimiento del aire en el recipiente, teniendo sólo en cuenta su compresión. Para el aire del cuello, sin embargo, lo que hicimos fue despreciar la energía de su compresión y tener sólo en cuenta la energía cinética de su movimiento.

Para deducir la ecuación diferencial de un péndulo se despreció la masa del hilo del cual pende. Para deducir la ecuación (14) de las oscilaciones eléctricas en un circuito se despreció la auto-inducción de los hilos y la resistencia de las bobinas. En general, para obtener la ecuación diferencial de un proceso físico se deben despreciar siempre ciertos factores e idealizar otros. En vista de esto, A. A. Andronov llamó la atención sobre el hecho de que a la física le interesan especialmente las ecuaciones diferenciales cuyas soluciones no varían mucho al modificar en una cantidad arbitrariamente pequeña las propias ecuaciones. Tales ecuaciones diferenciales se llaman «poco sensibles», o estructuralmente estables. Estas ecuaciones merecen un estudio particularmente completo.

Debe notarse que en las investigaciones físicas no sólo están imprecisamente definidas las ecuaciones diferenciales que describen las leyes de variación de dichas magnitudes físicas, sino que incluso el número de estas magnitudes sólo está definido aproximadamente. Estrictamente hablando no existen en realidad cuerpos rígidos. Así, al estudiar las oscilaciones de un péndulo, hay que tener en cuenta la deformación del hilo del cual cuelga y la deformación del propio cuerpo rígido, que nosotros hemos considerado aproximadamente como un punto material. Del mismo modo, para estudiar las oscilaciones de una masa sujeta por muelles debemos considerar las masas

de las espiras de los muelles. Pero en estos ejemplos puede demostrarse fácilmente que el carácter del movimiento de las diferentes partículas que forman el péndulo y los muelles influyen bien poco sobre el carácter de la oscilación. Si quisiéramos tener en cuenta esta influencia, el problema sería tan complicado que no se podría resolver con una aproximación conveniente. La solución no tendría mayor sentido físico que la dada en el §1 sin considerar estas influencias. La idealización inteligente de un problema es siempre inevitable. Para describir un proceso es necesario tener en cuenta los rasgos esenciales del mismo, pero no todos y cada uno de sus aspectos sin excepción. Esto no sólo complicaría en exceso el problema, sino que en muchos casos resultaría imposible encontrar su solución. El problema fundamental de la física o la mecánica, en la investigación de cualquier fenómeno, es encontrar el menor número de magnitudes, que, con una exactitud suficiente, describan el estado del fenómeno en un instante dado, y luego encontrar la ecuación diferencial más sencilla que constituya una buena descripción de las leyes que gobiernan los cambios de dichas magnitudes. Este problema es a menudo muy difícil. Qué rasgos son esenciales y cuáles no es una cuestión que en último término sólo puede decidirse por la experiencia. Únicamente comparando las soluciones proporcionadas por un razonamiento idealizado con los resultados del experimento se puede juzgar si la idealización es válida.

El problema matemático de la posibilidad de disminuir el número de variables puede formularse, en uno de los casos más sencillos y característicos, de la forma siguiente.

Supongamos para empezar que vamos a caracterizar el estado de un sistema físico en el instante t por dos magnitudes $x_1(t)$ y $x_2(t)$. Supongamos que las ecuaciones diferenciales que expresan su ritmo de variación son de la forma

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(t, x_1, x_2), \\ \epsilon \frac{dx_2}{dt} &= f_2(t, x_1, x_2). \end{aligned} \quad (28)$$

En la segunda ecuación el coeficiente de la derivada es un parámetro constante pequeño ϵ . Si hacemos $\epsilon = 0$, la segunda de las ecuaciones (28) no será una ecuación diferencial. Tendrá entonces la forma

$$f_2(t, x_1, x_2) = 0.$$

A partir de esta ecuación definimos x_2 como función de t y x_1 , y la sustituimos en la primera de las ecuaciones (28). Tenemos entonces la ecuación diferencial

$$\frac{dx_1}{dt} = F(t, x_1)$$

de la única variable x_1 . De este modo el número de parámetros que aparecen en el problema se reduce a uno. Nos preguntamos ahora en qué condiciones será pequeño el error introducido al tomar $\epsilon = 0$. Naturalmente, puede suceder que cuando $\epsilon \rightarrow 0$ el valor dx_2/dt crezca ilimitadamente, por lo que el segundo miembro de la segunda de las ecuaciones (28) no tenderá a cero cuando $\epsilon \rightarrow 0$.

§4. Interpretación geométrica del problema de integrar ecuaciones diferenciales; generalización del problema

Por sencillez sólo consideraremos en principio una ecuación diferencial de primer orden con una función incógnita

$$\frac{dy}{dx} = f(x, y) \quad (29)$$

donde la función $f(x, y)$ se supone definida sobre un dominio G del plano (x, y) . Esta ecuación determina, en cada punto del dominio, la pendiente de la tangente a la gráfica de una solución de la ecuación (29) en ese punto. Si en cada punto (x, y) del dominio G indicamos por medio de un segmento la dirección de la tangente tal y como viene determinada por el valor de $f(x, y)$ en ese punto, obtenemos un campo de direcciones. Entonces el problema de encontrar una solución de la ecuación diferencial (29) con la condición inicial $y(x_0) = y_0$ puede formularse así: encontrar en el dominio G una curva $y = \phi(x)$ que pase por el punto $M_0(x_0, y_0)$ y que en cada uno de sus puntos tenga una tangente cuya pendiente venga dada por la ecuación (29) o, dicho brevemente, que tenga en cada uno de sus puntos una dirección previamente asignada.

Desde un punto de vista geométrico este planteamiento del problema tiene dos imprecisiones:

1. Pedir que la pendiente de la tangente en un punto dado (x, y) del dominio G sea igual a $f(x, y)$ excluye automáticamente las tangentes paralelas a Oy , puesto que generalmente consideramos sólo

magnitudes finitas; en particular, se supone que la función $f(x, y)$ del segundo miembro de la ecuación (29) toma sólo valores finitos.

2. Al considerar sólo curvas que son gráficas de funciones de x excluimos también aquellas curvas que cortan más de una vez a una recta perpendicular al eje Ox , puesto que sólo se consideran funciones uniformes; en particular, toda solución de una ecuación diferencial se supone función uniforme de x .

Así, pues, generalicemos en cierta medida el aserto anterior, referente al problema de encontrar una solución a la ecuación diferencial (29), y permitamos que la tangente sea en algunos puntos paralela al eje Oy . En estos puntos, donde la pendiente de la tangente respecto al eje Ox no está definida, tomaremos la pendiente con relación al eje Oy . En otras palabras, consideraremos, junto con la ecuación diferencial (29), la ecuación

$$\frac{dx}{dy} = f_1(x, y), \quad (29')$$

donde $f_1(x, y) = 1/f(x, y)$, si $f(x, y) \neq 0$, utilizando la segunda ecuación cuando la primera carezca de sentido. El problema de integrar las ecuaciones diferenciales (29) y (29') es entonces el siguiente: Encontrar en el dominio G todas las curvas que tengan en cada punto la tangente definida por estas ecuaciones. Estas curvas se llamarán curvas integrales (líneas integrales) de las ecuaciones (29) y (29'), o del campo de tangentes definido por estas ecuaciones. En lugar del plural «ecuaciones (29), (29')», utilizaremos a menudo el singular «ecuación (29), (29')». Es claro que la gráfica de cualquier solución de la ecuación (29) será también una curva integral de la ecuación (29), (29'). Pero no toda curva integral de la ecuación (29), (29') será la gráfica de una solución de la ecuación (29). Este caso ocurrirá, por ejemplo, si alguna perpendicular al eje Ox corta a esta curva en más de un punto.

En lo que sigue, si se puede demostrar que

$$f(x, y) = \frac{M(x, y)}{N(x, y)},$$

escribiremos sólo la ecuación

$$\frac{dy}{dx} = \frac{M(x, y)}{N(x, y)},$$

omitiendo esta otra:

$$\frac{dx}{dy} = \frac{N(x, y)}{M(x, y)}.$$

A veces en lugar de estas ecuaciones introduciremos un parámetro t , y escribiremos el sistema de ecuaciones

$$\frac{dx}{dt} = N(x, y), \quad \frac{dy}{dt} = M(x, y),$$

donde x e y se consideran funciones de t .

Ejemplo 1

La ecuación

$$\frac{dy}{dx} = \frac{y}{x} \quad (30)$$

define un campo de tangentes en todo punto, excepto en el origen. Este campo de tangentes está esquematizado en la figura 7. Todas las tangentes dadas por la ecuación (30) pasan por el origen.

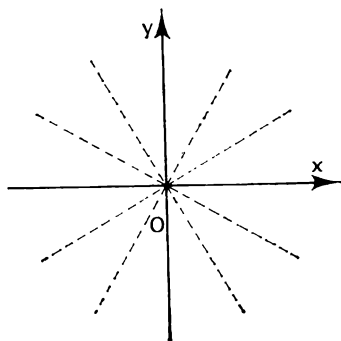


FIG. 7

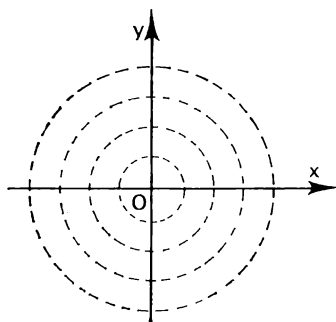


FIG. 8

Es claro que para todo k la función

$$y = kx \quad (31)$$

es una solución de la ecuación (30). La colección de todas las curvas integrales de esta ecuación está entonces definida por la relación

$$ax - by = 0, \quad (32)$$

donde a y b son constantes arbitrarias no simultáneamente nulas. El eje Oy es una curva integral de la ecuación (30), pero no es la gráfica de una de sus soluciones.

Puesto que la ecuación (30) no define un campo de tangentes en el origen, las curvas (31) y (32) son, estrictamente hablando, curvas integrales en todo punto, excepto en el origen. Sería, pues, más correcto decir que las curvas integrales de la ecuación (30) no son rectas que pasan por el origen sino semirectas que parten de él.

Ejemplo 2

La ecuación

$$\frac{dy}{dx} = -\frac{x}{y} \quad (33)$$

define un campo de tangentes en todo punto, excepto en el origen, como se ve en la figura 8. Las tangentes definidas en un punto dado (x, y) por las ecuaciones (30) y (33) son mutuamente perpendiculares. Es claro que todas las circunferencias centradas en el origen serán curvas integrales de la ecuación (33). Sin embargo, las soluciones de esta ecuación serán las funciones

$$y = +\sqrt{R^2 - x^2}, \quad y = -\sqrt{R^2 - x^2}, \quad -R \leq x \leq R.$$

Para abreviar, en lo que sigue diremos a veces «una solución pasa por el punto (x, y) » en lugar de la expresión más precisa «la gráfica de una solución pasa por el punto (x, y) ».

§5. Existencia y unicidad de la solución de una ecuación diferencial; solución aproximada de ecuaciones

El problema de la existencia y unicidad de la solución de una ecuación diferencial

Volvamos a la ecuación diferencial (17) de orden n . En general tiene infinitas soluciones, y, para poder escoger una entre todas las posibles, es necesario añadir condiciones suplementarias a la ecuación, en número igual al orden n de la ecuación. Tales condiciones pueden ser de carácter muy variado, dependiendo del sentido físico, mecánico, etc., del problema. Por ejemplo, si se trata de investigar el movimiento de un sistema mecánico que comienza en un estado

inicial dado, las condiciones suplementarias se referirán a un valor específico (inicial) de la variable independiente y se llamarán condiciones iniciales del problema. Pero, si se trata de obtener la forma de la curva que adopta un cable en un puente colgante o la de una viga cargada descansando sobre soportes en cada extremo, se darán condiciones correspondientes a distintos valores de la variable independiente en los extremos del cable o en los puntos soporte de la viga. Y podríamos citar otros muchos ejemplos ilustrativos de la variedad de condiciones que se pueden dar en relación con las ecuaciones diferenciales.

Supongamos que las condiciones suplementarias ya están dadas y se pide encontrar una solución de la ecuación (17) que las satisfaga. Lo primero a considerar es la existencia de tal solución. Sucede con frecuencia que dicha existencia no se puede garantizar de antemano. Supongamos, por ejemplo, que la ecuación (17) describe el modo de operar de un aparato físico y se desea saber si en él tiene lugar o no un movimiento periódico. Las condiciones suplementarias serán entonces condiciones para la repetición periódica del estado inicial en el aparato, y no se puede decir por adelantado si existirá o no una solución que las satisfaga.

En cualquier caso el estudio de los problemas de existencia y unicidad de una solución indica qué condiciones pueden ser verificadas por una ecuación diferencial dada y cuáles de estas condiciones definirán la solución de modo único. Pero la determinación de tales condiciones y la demostración de la existencia y unicidad de la solución de una ecuación diferencial correspondiente a un cierto problema físico encierra también un gran valor para la física teórica. Demuestra que las suposiciones adoptadas al plantear la descripción matemática del suceso físico son, por una parte, mutuamente compatibles, y por otra, constituyen una descripción completa del suceso.

Los métodos de investigación de los problemas de existencia son numerosos, pero de ellos son especialmente importantes los llamados métodos directos. La demostración de la existencia de la solución buscada se obtiene por construcción de soluciones aproximadas, demostrando luego que convergen a la solución exacta del problema. Estos métodos no sólo establecen la existencia de una solución exacta, sino que también proporcionan un medio (de hecho el más importante) de aproximarla con cualquier precisión deseada.

En el resto de esta sección consideraremos, para concretar, un problema con datos iniciales, con el cual ilustraremos la esencia del método de Euler y del de las aproximaciones sucesivas.

Método de Euler de las quebradas

Consideremos en un dominio G del plano Oxy la ecuación diferencial

$$\frac{dy}{dx} = f(x, y). \quad (34)$$

Como ya hemos indicado, la ecuación (34) define en G un campo de tangentes. Elijamos cualquier punto (x_0, y_0) de G . Por él pasará una recta L_0 de pendiente $f(x_0, y_0)$. Sobre la recta L_0 elegimos un punto (x_1, y_1) , suficientemente próximo a (x_0, y_0) ; en la figura 9 este punto está indicado con el número 1.

Por el punto (x_1, y_1) trazamos la recta L_1 de pendiente $f(x_1, y_1)$, y sobre ella marcamos el punto (x_2, y_2) ; en la figura este punto viene indicado por el número 2. Luego, sobre la línea recta L_2 correspondiente al

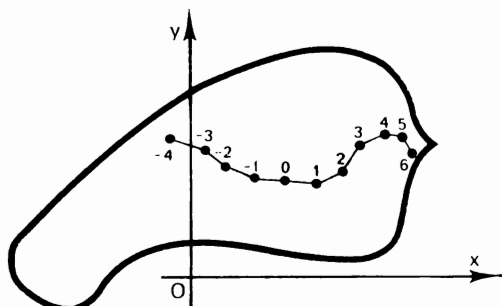


FIG. 9

punto (x_2, y_2) , marcamos el punto (x_3, y_3) , y continuamos del mismo modo con $x_0 < x_1 < x_2 < x_3 < \dots$. Se supone, naturalmente, que los puntos (x_0, y_0) , (x_1, y_1) , (x_2, y_2) , $\dots$, están todos ellos en el dominio G . La línea quebrada que une estos puntos se llama quebrada de Euler. También puede construirse una quebrada de Euler en la dirección de las x decrecientes; los correspondientes vértices en la figura se notan por -1 , -2 , -3 .

Es razonable esperar que toda quebrada de Euler que pase por el punto (x_0, y_0) y esté formada por segmentos suficientemente pequeños proporcione una representación de una curva integral l que pase por el punto (x_0, y_0) ; como también es lógico que al decrecer las longitudes de los lados de la quebrada, es decir, cuando la longitud del lado mayor tienda a cero, aquélla se aproxime a la curva integral.

Hasta ahora, naturalmente, se ha supuesto que la curva integral existe. De hecho no es difícil demostrar que si la función $f(x, y)$ es continua en el dominio G , se puede encontrar una sucesión infinita

de quebradas de Euler, cuyos lados mayores tienden a cero, que converge a una curva integral I . Sin embargo, en general no se puede demostrar la unicidad: pueden existir sucesiones distintas de quebradas de Euler que converjan a curvas integrales diferentes pasando por el mismo punto (x_0, y_0) . M. A. Laurentiev ha dado un ejemplo de ecuación diferencial de la forma (29) con una función continua $f(x, y)$ tal que en todo entorno de un punto cualquiera P del dominio G pase no una, sino al menos dos curvas integrales. Para que por todo punto del dominio G sólo pase una única curva integral, es necesario imponer a la función $f(x, y)$ ciertas condiciones además de la de continuidad. Es suficiente, por ejemplo, suponer que la función $f(x, y)$ sea continua y tenga derivada acotada con respecto a y en todo el dominio G . En este caso puede demostrarse que por todo punto de G pasa una y sólo una curva integral y que toda sucesión de quebradas de Euler que pasen por el punto (x_0, y_0) converge uniformemente a esta única curva integral, cuando la longitud del lado mayor de la quebrada tiende a cero. Así, para lados suficientemente pequeños, la quebrada de Euler puede considerarse como una aproximación a la curva integral de la ecuación (34).

De lo anterior puede verse que las quebradas de Euler se construyen de modo que pequeños trozos de las curvas integrales se reemplazan por segmentos de rectas tangentes a estas curvas integrales. En la práctica, las aproximaciones a las curvas integrales de la ecuación diferencial (34) constan a menudo, no de segmentos de recta tangentes a las curvas integrales, sino de segmentos parabólicos que tienen un orden superior de contacto con la curva integral. De este modo se puede encontrar una solución aproximada con el mismo grado de aproximación que antes, pero en un número de pasos menor (con menor número de eslabones en la curva aproximada). Los coeficientes de la ecuación para la parábola (de orden superior)

$$y = a_0 + a_1(x - x_k) + a_2(x - x_k)^2 + \cdots + a_n(x - x_k)^n, \quad (35)$$

que tiene en el punto (x_k, y_k) un orden n de contacto con las curvas integrales de la ecuación (34) en este punto, vienen dados por las fórmulas siguientes:

$$a_0 = y_k \quad (36)$$

$$a_1 = \left(\frac{dy}{dx} \right)_{x=x_k} = f(x_k, y_k), \quad (36')$$

$$\begin{aligned}
 2a_2 &= \left(\frac{d^2 y}{dx^2} \right)_{x=x_k} = \left[\frac{df(x, y)}{dx} \right]_{x=x_k} = \\
 &= f'_x(x_k, y_k) + f'_y(x_k, y_k) \left(\frac{dy}{dx} \right)_{x=x_k} = \\
 &= f'_x(x_k, y_k) + f'_y(x_k, y_k) f(x_k, y_k), \quad (36'')
 \end{aligned}$$

$$\begin{aligned}
 6a_3 &= \left(\frac{d^3 y}{dx^3} \right)_{x=x_k} = \left\{ \frac{d}{dx} \left[f'_x(x, y(x)) + f'_y(x, y(x)) f(x, y(x)) \right] \right\}_{x=x_k} = \\
 &= f''_{xx}(x_k, y_k) + 2f''_{xy}(x_k, y_k) f(x_k, y_k) + \\
 &+ f''_{yy}(x_k, y_k) f^2(x_k, y_k) + f'^2_{y^2}(x_k, y_k) f(x_k, y_k) + \\
 &+ f'_y(x_k, y_k) f'_x(x_k, y_k). \quad (36''')
 \end{aligned}$$

El polinomio (35) sólo se necesita para obtener su valor en $x = x_{k+1}$. Los valores propiamente dichos de los coeficientes $a_0, a_1, a_2, \dots, a_n$ no son necesarios. Hay muchos modos de calcular el valor para $x = x_{k+1}$ del polinomio (35), cuyos coeficientes vienen dados por la fórmula (36), sin hallar efectivamente los coeficientes $a_0, a_1, \dots, a_n$.

Existen otros métodos de aproximación de la solución de la ecuación diferencial (34) basados en otras ideas. Un método conveniente lo desarrolló A. N. Krylov (1863-1945).

El método de las aproximaciones sucesivas

Describamos ahora otro método de aproximaciones sucesivas, que es tan utilizado como el de las quebradas de Euler. Supongamos de nuevo que se pide encontrar una solución $y(x)$ de la ecuación diferencial (34) que satisfaga la condición inicial

$$y(x_0) = y_0.$$

Como primera aproximación de la función $y(x)$ tomamos una función arbitraria $y_0(x)$. Por razones de sencillez supondremos que también satisface la condición inicial, aunque esto no es necesario. Sustituimos ahora la función y del segundo miembro $f(x, y)$ de la ecuación por $y_0(x)$, y construimos una primera aproximación y_1 a la solución y con las siguientes condiciones:

$$\frac{dy_1}{dx} = f \left[x, y_0(x) \right], \quad y_1(x_0) = y_0.$$

Puesto que la función del segundo miembro de la primera de estas ecuaciones es conocida, puede encontrarse la función $y_1(x)$ por integración

$$y_1(x) = y_0 + \int_{x_0}^x f \left[t, y_0(t) \right] dt.$$

Cabe esperar que $y_1(x)$ difiera de la solución $y(x)$ en menos que $y_0(x)$, puesto que para obtener $y_1(x)$ se ha utilizado la propia ecuación diferencial, que probablemente introducirá una corrección en la primera aproximación. Asimismo, cabe pensar que, si se mejora la primera aproximación $y_1(x)$ del mismo modo, entonces la aproximación segunda

$$y_2(x) = y_0 + \int_{x_0}^x f \left[t, y_1(t) \right] dt$$

estará todavía más próxima a la solución buscada.

Supongamos que este proceso de mejora continúa indefinidamente y que se ha obtenido la sucesión de aproximaciones

$$y_0(x), y_1(x), \dots, y_n(x), \dots$$

¿Convergerá esta sucesión a la solución $y(x)$?

Estudios más precisos demuestran que, si $f(x, y)$ es continua y f'_y acotada en el dominio G , las funciones $y_n(x)$ convergerán en efecto a la solución exacta $y(x)$, para x suficientemente próximo a x_0 ; y que, si detenemos el cálculo después de un número suficiente de pasos, se puede determinar la solución $y(x)$ con cualquier grado deseado de precisión.

Exactamente del mismo modo que para las curvas integrales de la ecuación (34), se pueden obtener también aproximaciones a las curvas integrales de un sistema de dos o más ecuaciones diferenciales de primer orden. En esencia la condición necesaria es poder despejar en estas ecuaciones las derivadas de las funciones desconocidas. Por ejemplo, supongamos que se da el sistema

$$\frac{dy}{dx} = f_1(x, y, z), \quad \frac{dz}{dx} = f_2(x, y, z). \quad (37)$$

Suponiendo que los segundos miembros de estas ecuaciones son continuos y tienen derivadas acotadas respecto a y y z en algún dominio G del espacio, puede demostrarse que, en estas condiciones,

por cada punto (x_0, y_0, z_0) del dominio G , en el que están definidos los segundos miembros de las ecuaciones (37), pasa una y sólo una curva integral,

$$y = \phi(x), \quad z = \psi(x),$$

del sistema (37). Las funciones $f_1(x, y, z)$ y $f_2(x, y, z)$ dan la dirección, en el punto (x, y, z) , de la tangente a la curva integral que pasa por este punto. Para encontrar las funciones $\phi(x)$ y $\psi(x)$ aproximadamente se puede aplicar el método de la quebrada de Euler u otros métodos parecidos a los que aplicamos a la ecuación (34).

El proceso del cálculo aproximado de la solución de las ecuaciones diferenciales ordinarias con condiciones iniciales puede ser realizado por computadores. Hay máquinas electrónicas que trabajan tan rápidamente que si, por ejemplo, están programadas para calcular la trayectoria de un proyectil, esta trayectoria puede encontrarse en un espacio de tiempo más corto que el que tarde el proyectil en alcanzar el blanco.

Conexión entre las ecuaciones diferenciales de distintos órdenes y los sistemas de ecuaciones de primer orden

Un sistema de ecuaciones diferenciales ordinarias en el que se haya despejado la derivada de mayor orden de cada una de las funciones incógnitas, puede reducirse en general, mediante la introducción de nuevas funciones desconocidas, a un sistema de ecuaciones de primer orden, en las cuales están despejadas todas las derivadas. Por ejemplo, consideremos la ecuación diferencial

$$\frac{d^2y}{dx^2} = f\left(x, y, \frac{dy}{dx}\right). \quad (38)$$

Pongamos

$$\frac{dy}{dx} = z. \quad (39)$$

La ecuación (38) puede escribirse entonces en la forma

$$\frac{dz}{dx} = f(x, y, z). \quad (40)$$

Por tanto, toda solución de la ecuación (38) va ligada a una solución del sistema formado por las ecuaciones (39) y (40). Es fácil demostrar que a toda solución del sistema de ecuaciones (39) y (40) corresponde una solución de la ecuación (38).

Ecuaciones que no contienen explícitamente la variable independiente

Los problemas del péndulo, del resonador acústico de Helmholtz, de un circuito eléctrico sencillo o de un generador electrónico, considerados en el §1, conducen a ecuaciones diferenciales en las que la variable independiente (tiempo) no aparece explícitamente. Mencionamos aquí las ecuaciones de este tipo porque las correspondientes ecuaciones diferenciales de segundo orden pueden reducirse en todos los casos a una única ecuación diferencial de primer orden, más que a un sistema de ecuaciones de primer orden, tal como se vio en la sección anterior para la ecuación general de segundo orden. Esta reducción simplifica grandemente su estudio.

Consideremos entonces una ecuación diferencial de segundo orden que no contiene la variable t en forma explícita

$$F\left(x, \frac{dx}{dt}, \frac{d^2x}{dt^2}\right) = 0. \quad (41)$$

Hagamos

$$\frac{dx}{dt} = y \quad (42)$$

y consideremos y como función de x , con lo cual

$$\frac{d^2x}{dt^2} = \frac{d}{dt} \left(\frac{dx}{dt} \right) = \frac{dy}{dt} = \frac{dy}{dx} \frac{dx}{dt} = y \frac{dy}{dx}.$$

La ecuación (41) puede escribirse entonces en la forma

$$F\left(x, y, y \frac{dy}{dx}\right) = 0. \quad (43)$$

De este modo, cada solución de la ecuación (41) va ligada a una solución única de la ecuación (43). Y a cada una de las soluciones $y = \phi(x)$ de la ecuación (43) corresponden infinitas soluciones de la ecuación (41). Estas soluciones pueden encontrarse integrando la ecuación

$$\frac{dx}{dt} = \phi(x), \quad (44)$$

donde x se considera función de t .

Es claro que, si esta ecuación se satisface para una función $x = x(t)$, entonces también se satisfará para cualquier función de la forma $x(t + t_0)$, siendo t_0 una constante arbitraria.

Puede suceder que no toda curva integral de la ecuación (43) sea la gráfica de una única función de x . Esto sucederá, por ejemplo, si la curva es cerrada. En este caso la curva integral de la ecuación (43) debe ser dividida en un cierto número de trozos, cada uno de los cuales sea la gráfica de una función de x . Para cada uno de estos trozos tendremos que encontrar una integral de la ecuación (44).

Los valores de x y de dx/dt que caracterizan en cada instante el estado del sistema físico correspondiente a la ecuación (41) se llaman *fases* del sistema, y el plano Oxy se llama *plano de las fases* de la ecuación (41). Toda solución $x = x(t)$ de esta ecuación va ligada a la curva

$$x = x(t), \quad y = x'(t)$$

en el plano Oxy ; t se considera aquí como parámetro. Recíprocamente, a toda curva integral $y = \phi(x)$ de la ecuación (43) en el plano Oxy le corresponde un conjunto infinito de soluciones de la forma $x = x(t + t_0)$ para la ecuación (41); t_0 es aquí una constante arbitraria. La información sobre el comportamiento de las curvas integrales de la ecuación (43) en el plano se transforma fácilmente en información sobre el carácter de las posibles soluciones de la ecuación (41). Toda curva integral cerrada de la ecuación (43) corresponde, por ejemplo, a una solución periódica de la ecuación (41). Si aplicamos a la ecuación (6) la transformación (42), se obtiene

$$\frac{dy}{dx} = \frac{-ay - bx}{my}. \quad (45)$$

Poniendo $v = x$ y $dv/dt = y$ en la ecuación (16), se obtiene de igual modo

$$L \frac{dy}{dx} = \frac{-[R - M(a_1 + 2a_2x + 3a_3x^2)]y - x}{y}. \quad (46)$$

Igual que en cada instante el estado del sistema físico correspondiente a la ecuación de segundo orden (41) está caracterizado por

la dos magnitudes³ (fases) x e $y = dx/dt$, el estado de un sistema físico descrito por una ecuación de orden superior o por un sistema de ecuaciones diferenciales está caracterizado por un gran número de magnitudes (fases). En lugar de un plano fase, hablaremos entonces de un espacio de las fases.

§6. Puntos singulares

Sea $P(x, y)$ un punto interior del dominio G en el que se considera la ecuación diferencial

$$\frac{dy}{dx} = \frac{M(x, y)}{N(x, y)}. \quad (47)$$

Si existe un entorno R del punto P tal que por cada uno de sus puntos pase una, y sólo una, curva integral (47), entonces el punto P se llama *punto ordinario* de la ecuación (47). Si no existe tal entorno, el punto P se llama *punto singular* de la ecuación. El estudio de los puntos singulares es muy importante en la teoría cualitativa de las ecuaciones diferenciales, teoría que se considerará en la sección siguiente.

Particularmente importantes son los llamados puntos *singulares aislados*, es decir, puntos singulares tales que en cierto entorno alrededor de él no hay otros puntos singulares. En las aplicaciones se encuentran con frecuencia puntos singulares al estudiar las ecuaciones de la forma (47), donde $M(x, y)$ y $N(x, y)$ son funciones con derivadas parciales continuas hasta un cierto orden respecto a x e y . Para tales ecuaciones, todos los puntos interiores del dominio en los cuales $M(x, y) \neq 0$ ó $N(x, y) \neq 0$ son puntos ordinarios. Consideremos ahora cualquier punto interior (x_0, y_0) en el que $M(x, y) = N(x, y) = 0$. Para simplificar la notación supondremos que $x_0 = 0$ e $y_0 = 0$. Esto siempre puede hacerse trasladando el origen de coordenadas al punto (x_0, y_0) . Desarrollando $M(x, y)$ y $N(x, y)$ por la fórmula de Taylor en potencias de x e y hasta los términos de primer orden, se tiene, en un entorno del punto $(0, 0)$,

$$\frac{dy}{dx} = \frac{M'_x(0, 0)x + M'_y(0, 0)y + \phi_1(x, y)}{N'_x(0, 0)x + N'_y(0, 0)y + \phi_2(x, y)}, \quad (48)$$

³ Los valores de d^2x/dt^2 , d^3x/dt^3 , ..., en el mismo instante de tiempo están definidos por los valores de x y dx/dt de la ecuación (41) y de las ecuaciones obtenidas de (45) por derivación —cf. fórmula (36).

donde $\phi_1(x, y)$ y $\phi_2(x, y)$ son funciones de x e y para las cuales

$$\lim_{\substack{x \rightarrow 0 \\ y \rightarrow 0}} \frac{\phi_1(x, y)}{\sqrt{x^2 + y^2}} = 0 \quad \text{y} \quad \lim_{\substack{x \rightarrow 0 \\ y \rightarrow 0}} \frac{\phi_2(x, y)}{\sqrt{x^2 + y^2}} = 0.$$

Las ecuaciones (45) y (46) son de esta forma. La ecuación (45) no define dy/dx o dx/dy para $x = 0$ e $y = 0$. Si el determinante

$$\begin{vmatrix} M'_x(0, 0) & M'_y(0, 0) \\ N'_x(0, 0) & N'_y(0, 0) \end{vmatrix} \neq 0$$

entonces, cualquiera que sea el valor que se asigne a dy/dx en el origen, éste será un punto de discontinuidad para los valores de dy/dx y dx/dy , puesto que tienden a límites distintos según el modo de acercarse al origen. El origen es un punto singular de la ecuación diferencial.

Se ha demostrado que el carácter del comportamiento de las curvas integrales cerca de un punto singular aislado (en este caso el origen) no se ve afectado por el comportamiento de los términos $\phi_1(x, y)$ y $\phi_2(x, y)$ en el numerador y denominador siempre que la parte real de ambas raíces de la ecuación

$$\begin{vmatrix} \lambda - M'_y(0, 0) & -M'_x(0, 0) \\ -N'_y(0, 0) & \lambda - N'_x(0, 0) \end{vmatrix} = 0 \quad (49)$$

sea distinta de cero. Así, para formarse una idea de este comportamiento, se estudia el comportamiento de las curvas integrales de la ecuación

$$\frac{dy}{dx} = \frac{ax + by}{cx + dy} \quad (50)$$

cerca del origen, para las cuales el determinante

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0.$$

Observemos que la disposición de las curvas integrales en el entorno de un punto singular de una ecuación diferencial es de gran interés para muchos problemas de la mecánica, por ejemplo, en el estudio

de las trayectorias de los movimientos próximos a la posición de equilibrio.

Se ha demostrado que en todo punto del plano se pueden elegir unas coordenadas ξ, η , relacionadas con las x, y por las ecuaciones

$$\begin{aligned}x &= k_{11}\xi + k_{12}\eta, \\y &= k_{12}\xi + k_{22}\eta,\end{aligned}\tag{51}$$

donde las k_{ij} son números reales tales que la ecuación (50) se transforma en uno de los tres tipos siguientes:

$$1) \quad \frac{d\eta}{d\xi} = k \frac{\eta}{\xi}, \quad \text{donde} \quad k = \frac{\lambda_2}{\lambda_1}.$$

$$2) \quad \frac{d\eta}{d\xi} = \frac{\xi + \eta}{\xi}.$$

$$3) \quad \frac{d\eta}{d\xi} = \frac{\beta\xi + \alpha\eta}{\alpha\xi - \beta\eta}.$$

λ_1 y λ_2 son aquí las raíces de la ecuación

$$\begin{vmatrix} c - \lambda & d \\ a & b - \lambda \end{vmatrix} = 0.$$

Si estas raíces son reales y distintas, la ecuación (50) se transforma en la ecuación (52). Si son iguales, entonces la ecuación (50) se transforma, bien en la (52), bien en la (53), según sea $a^2 + d^2 = 0$ ó $a^2 + d^2 \neq 0$. Si las raíces de la ecuación (55) son complejas, $\lambda = \alpha \pm \beta i$, entonces la ecuación (51) se transforma en la (54).

Consideremos cada una de las ecuaciones (52), (53), (54). Para empezar, observemos lo siguiente.

Aunque los ejes Ox y Oy fuesen mutuamente perpendiculares, los ejes $O\xi$ y $O\eta$ no necesitan, en general, serlo. Pero, para simplificar los diagramas, supondremos que son perpendiculares. Además, en la transformación (51) las escalas de los ejes $O\xi$ y $O\eta$ pueden variar, y no ser las mismas que las elegidas para los ejes Ox y Oy . Pero para simplificar supondremos que las escalas son las mismas. Así, por ejemplo, en lugar de circunferencias concéntricas, como en la figura 8,

podría aparecer en general una familia de elipses semejantes y semejantemente situadas, con centro común en el origen.

Todas las curvas integrales de la ecuación (52) vienen dadas por una relación de la forma:

$$a\eta + b|\xi|^k = 0,$$

donde a y b son constantes arbitrarias.

Las curvas integrales de la ecuación (52) están representadas en la figura 10, donde se ha supuesto que $k > 1$. En este caso todas las curvas integrales excepto una, el eje $O\eta$, son tangentes en el origen

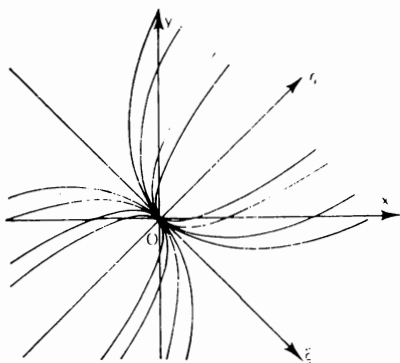


FIG. 10

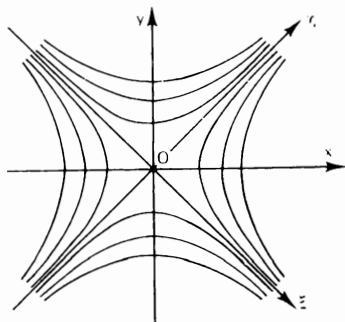


FIG. 11

al eje $O\xi$. El caso $0 < k < 1$ es el mismo que el caso $k > 1$ intercambiando ξ y η , es decir, sólo hay que intercambiar los papeles de los ejes ξ y η . Para $k = 1$, la ecuación (52) se convierte en la ecuación (30), cuyas curvas integrales se representaron en la figura 7.

En la figura 11 se da una representación de las curvas integrales de la ecuación (52) para $k < 0$. En este caso únicamente hay dos curvas integrales que pasan por el punto O : éstas son el eje $O\xi$ y el eje $O\eta$. Todas las restantes curvas integrales, después de aproximarse al origen hasta una cierta distancia mínima, se alejan de él. En este caso se dice que el punto O es un punto de *ensilladura*.

Todas las curvas integrales de la ecuación (53) vienen dadas por la ecuación

$$b\eta = \xi(a + b \ln |\xi|),$$

donde a y b son constantes arbitrarias. Esto se ilustra esquemáticamente en la figura 12; todas ellas son tangentes al eje $O\eta$ en el origen.

Si toda curva integral que penetra en cierto entorno del punto singular O pasa por este punto y tiene una dirección definida en él, es decir, tiene una tangente definida en el origen, como se ve en las figuras 10 y 12, entonces se dice que el punto O es un *nodo*.

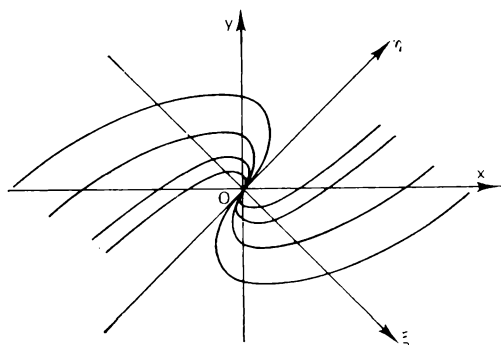


FIG. 12

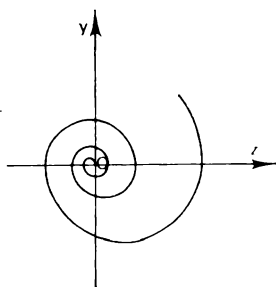


FIG. 13

La ecuación (54) resulta más fácil de integrar si se cambia a coordenadas polares ϱ y ϕ , poniendo

$$\xi = \varrho \cos \phi, \quad \eta = \varrho \sin \phi.$$

Esta ecuación se transforma entonces en la siguiente:

$$\frac{d\varrho}{d\phi} = k\varrho \quad \text{donde} \quad k = \frac{a}{\beta},$$

y por consiguiente,

$$\varrho = Ce^{k\phi}. \quad (56)$$

Si $K > 0$ todas las curvas integrales se aproximan al punto O según una espiral que rodea infinitas veces este punto cuando $\phi \rightarrow -\infty$ (Fig. 13). Si $k < 0$, entonces esto sucede para $k \rightarrow +\infty$. En estos casos, el punto O se llama *foco*. Sin embargo, si $k = 0$, la colección de curvas integrales de (56) consta de curvas con centro en el punto O . Generalmente, si algún entorno del punto O está completamente ocupado por curvas integrales cerradas, rodeando al mismo punto O , tal punto se llama *centro*.

Un centro puede transformarse fácilmente en foco, si al numerador y denominador del segundo miembro de la ecuación (54) añadimos un término de orden arbitrariamente alto; por tanto, en este caso el comportamiento de las curvas integrales cerca de un punto singular no viene dado por términos de primer orden.

La ecuación (55), correspondiente a la ecuación (45), es idéntica a la ecuación característica (19). Así las figuras 10 y 12 representan esquemáticamente el comportamiento en el plano fase Oxy de las curvas

$$\dot{x} = x(t), \quad \dot{y} = x'(t),$$

correspondientes a las soluciones de la ecuación (6) para λ_1 y λ_2 reales y del mismo signo; la figura 11 corresponde a λ_1 y λ_2 reales de signos opuestos, y las figuras 13 y 8 (el caso de un centro) corresponden a λ_1 y λ_2 complejos. Si las partes reales de λ_1 y λ_2 son negativas, entonces el punto $(x(t), y(t))$ tiende a cero para $t \rightarrow +\infty$; en este caso el punto $x = 0, y = 0$ corresponde a un equilibrio estable. Sin embargo, si la parte real de uno cualquiera de los números λ_1 y λ_2 es positiva, entonces en el punto $x = 0, y = 0$ no hay equilibrio estable.

§7. Teoría cualitativa de las ecuaciones diferenciales ordinarias

Una parte importante de la teoría general de las ecuaciones diferenciales ordinarias es la teoría cualitativa de las ecuaciones diferenciales, que surgió a finales del último siglo por necesidades de la mecánica y la astronomía.

En muchos problemas prácticos es necesario establecer el carácter de la solución de una ecuación diferencial que describe un proceso físico y determinar las propiedades de sus soluciones cuando la variable independiente varía, o se mueve, sobre un intervalo finito o infinito. Por ejemplo, en la mecánica celeste, que estudia el movimiento

de los cuerpos en el espacio exterior, es importante tener información sobre el comportamiento de las soluciones de las ecuaciones diferenciales que describen el movimiento de los planetas u otros cuerpos celestes en períodos no limitados de tiempo.

Como se vio anteriormente, sólo para unas cuantas ecuaciones particularmente sencillas se puede expresar la solución general en términos de integrales de funciones conocidas. De este modo surgió el problema de investigar las propiedades de las soluciones de una ecuación diferencial a partir de la ecuación misma. Puesto que la solución de una ecuación diferencial viene dada en forma de una curva en el plano o en el espacio, el problema consiste en investigar las propiedades de las curvas integrales, su distribución y su comportamiento en el entorno de los puntos singulares. Por ejemplo, ¿se encuentran en una parte acotada del plano, o tienen ramas que tienden a infinito? ¿Alguna de ellas es curva cerrada?, etc. La investigación de tales cuestiones constituye la teoría cualitativa de las ecuaciones diferenciales.

Los fundadores de la teoría cualitativa de las ecuaciones diferenciales son el matemático ruso A. M. Lyapunov y el matemático francés H. Poincaré.

En la sección precedente consideramos ya en detalle una de las importantes cuestiones de la teoría cualitativa, la distribución de las curvas integrales en un entorno de un punto singular. Volvamos ahora a algunas otras cuestiones básicas de la teoría cualitativa.

Estabilidad

En los ejemplos considerados al comienzo del capítulo, la cuestión de la estabilidad o inestabilidad del equilibrio de un sistema se resolvió fácilmente por consideraciones físicas, sin estudiar las ecuaciones diferenciales. Así, en el ejemplo 3 es obvio que si, en la posición de equilibrio OA , el péndulo se mueve, bajo la acción de una fuerza externa, hasta una posición próxima OA' , es decir, si se hace un pequeño cambio en las condiciones iniciales, entonces el movimiento del péndulo no diferirá grandemente de la posición de equilibrio, y esta desviación será tanto más pequeña cuanto menores sean las desviaciones originales OA' , es decir, en este caso la posición de equilibrio será estable.

Para otros casos más complicados, la cuestión de la estabilidad de la posición de equilibrio es considerablemente más complicada y sólo puede resolverse estudiando las correspondientes ecuaciones diferenciales. El problema de la estabilidad del equilibrio está íntima-

mente relacionado con la cuestión de la estabilidad del movimiento. A. M. Lyapunov estableció en este campo algunos resultados fundamentales.

Consideremos un proceso físico descrito por el sistema de ecuaciones

$$\begin{aligned}\frac{dx}{dt} &= f_1(x, y, t), \\ \frac{dy}{dt} &= f_2(x, y, t).\end{aligned}\tag{57}$$

Por sencillez, sólo consideramos un sistema de dos ecuaciones diferenciales, aunque las conclusiones siguen siendo válidas para sistemas con mayor número de ecuaciones. Cada una de las soluciones particulares del sistema (57), que constará de dos funciones $x(t)$ e $y(t)$, se llamará a veces movimiento, siguiendo a Lyapunov. Supondremos que $f_1(x, y, t)$ y $f_2(x, y, t)$ tienen derivadas parciales continuas. Se ha demostrado que en este caso la solución del sistema de ecuaciones diferenciales (57) está unívocamente definido si se dan en cualquier instante de tiempo $t = t_0$ los valores iniciales $x(t_0) = x_0$ e $y(t_0) = y_0$.

Notaremos por $x(t, x_0, y_0)$ e $y(t, x_0, y_0)$ la solución del sistema de ecuaciones (57) que satisfaga las condiciones iniciales

$$x = x_0 \text{ e } y = y_0 \quad \text{para } t = t_0.$$

Una solución $x(t, x_0, y_0)$, $y(t, x_0, y_0)$ se llama *estable en el sentido de Lyapunov* si, para todo $t > t_0$, las funciones $x(t, x_0, y_0)$ e $y(t, x_0, y_0)$ experimentan cambios pequeños para modificaciones suficientemente pequeñas en los valores iniciales de x_0 e y_0 .

Con más precisión, una solución es estable en el sentido de Lyapunov si las diferencias

$$\begin{aligned}&|x(t, x_0 + \delta_1, y_0 + \delta_2) - x(t, x_0, y_0)|, \\ &|y(t, x_0 + \delta_1, y_0 + \delta_2) - y(t, x_0, y_0)|\end{aligned}\tag{58}$$

pueden hacerse menores que cualquier número dado ϵ para todo $t > t_0$, siempre que los números δ_1 y δ_2 se tomen suficientemente pequeños en valor absoluto.

Todo movimiento que no es estable en el sentido de Lyapunov se llama *inestable*.

En sus investigaciones, Lyapunov llamó al movimiento $x(t, x_0, y_0)$ e $y(t, x_0, y_0)$ no perturbado; y perturbado, al movimiento $x(t, x_0 + \delta_1, y_0 + \delta_2)$, $y(t, x_0 + \delta_1, y_0 + \delta_2)$ con condiciones iniciales próximas. De este modo la estabilidad en el sentido de Lyapunov para un movimiento no perturbado significa que para todo $t > t_0$ el movimiento perturbado sólo debe diferir un poco del no perturbado.

La estabilidad del equilibrio es un caso especial de la estabilidad del movimiento, correspondiendo al caso en el que el movimiento no perturbado es

$$x(t, x_0, y_0) \equiv 0 \quad \text{e} \quad y(t, x_0, y_0) \equiv 0.$$

Recíprocamente, la cuestión de la estabilidad de cualquier movimiento $x = \phi_1(t)$ e $y = \phi_2(t)$ del sistema (57) puede ser reducida a la de la estabilidad del equilibrio de un cierto sistema de ecuaciones diferenciales. Para ello se reemplazan las funciones desconocidas $x(t)$ e $y(t)$ en el sistema (57) por las nuevas funciones incógnitas

$$\xi = x - \phi_1(t) \quad \text{y} \quad \eta = y - \phi_2(t). \quad (59)$$

En el sistema (57), transformado de este modo, el movimiento $x = \phi_1(t)$ e $y = \phi_2(t)$ corresponderá al movimiento $\xi \equiv 0$ y $\eta \equiv 0$, es decir, a la posición de equilibrio. En lo que sigue se supondrá siempre que se ha efectuado la transformación (59), por lo que basta considerar la estabilidad en el sentido de Lyapunov para la solución $x = 0, y = 0$.

La condición de estabilidad en el sentido de Lyapunov significa ahora que, para δ_1 y δ_2 suficientemente pequeños y $t > t_0$, la trayectoria en el plano Oxy de un movimiento perturbado no pasa fuera del cuadrado de lados de longitud 2 paralelos a los ejes de coordenadas y de centro el punto $x = 0, y = 0$.

Particularmente interesantes son aquellos casos en los cuales, sin conocer una integral del sistema (57), se puede, sin embargo, sacar conclusiones sobre la estabilidad o inestabilidad de un movimiento. La estabilidad es una cuestión práctica muy importante en el movimiento de proyectiles o de aeroplanos; y la estabilidad de las órbitas es importante en la mecánica celeste, donde los movimientos de los planetas y otros cuerpos en el espacio exterior conducen a este tipo de investigaciones.

Supongamos que las funciones $f_1(x, y, t)$ y $f_2(x, y, t)$ pueden representarse en la forma

$$\begin{aligned} f_1(x, y, t) &= a_{11}x + a_{12}y + R_1(x, y, t), \\ f_2(x, y, t) &= a_{21}x + a_{22}y + R_2(x, y, t), \end{aligned} \quad (60)$$

donde las a_{ij} son constantes, y $R_1(x, y, t)$ y $R_2(x, y, t)$ son funciones de x, y, t tales que

$$|R_1(x, y, t)| \leq M(x^2 + y^2) \quad \text{y} \quad |R_2(x, y, t)| \leq M(x^2 + y^2), \quad (61)$$

donde M es una constante positiva.

Si en el sistema (57) sustituimos las ecuaciones (60), despreciando $R_1(x, y, t)$ y $R_2(x, y, t)$, obtenemos un sistema de ecuaciones diferenciales con coeficientes constantes

$$\begin{aligned} \frac{dx}{dt} &= a_{11}x + a_{12}y \\ \frac{dy}{dt} &= a_{21}x + a_{22}y \end{aligned} \quad (62)$$

que se llama *sistema de primera aproximación del sistema no lineal* (57).

Antes de Lyapunov, los investigadores se limitaban a estudiar la estabilidad de la primera aproximación, creyendo que los resultados obtenidos se podrían utilizar para estudiar la estabilidad del sistema básico no lineal (57). Lyapunov fue el primero en demostrar que, en general, esta conclusión es falsa. Por otra parte, dio una serie de condiciones muy amplias bajo las cuales la estabilidad de un sistema no lineal se resuelve completamente con la primera aproximación. Una de estas condiciones es la siguiente. Si las partes reales de las dos raíces de la ecuación

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0$$

son negativas, y las funciones $R_1(x, y, t)$ y $R_2(x, y, t)$ verifican la condición (61), entonces la solución $x(t) \equiv 0$, $y(t) \equiv 0$ es estable en el sentido de Lyapunov. Si la parte real de sólo una de las raíces es positiva, entonces la solución $x(t) \equiv 0$, $y(t) \equiv 0$ de una ecuación que satisfaga las condiciones (61) es inestable. Lyapunov dio también

una serie de condiciones suficientes para la estabilidad y la inestabilidad de un movimiento ⁴.

Si los segundos miembros de las ecuaciones (57) no dependen de t , dividiendo la primera ecuación del sistema (57) por la segunda se obtiene

$$\frac{dy}{dx} = \frac{f_1(x, y)}{f_2(x, y)}. \quad (63)$$

El origen será un punto singular de esta ecuación. En el caso de la estabilidad del equilibrio, este punto puede ser un foco, un nodo o un centro, pero no puede ser un punto de ensilladura.

Puede determinarse así el carácter de un punto singular a partir de la estabilidad o inestabilidad de la posición de equilibrio.

Comportamiento global de las curvas integrales

A veces es importante obtener una representación esquemática del comportamiento «global» de las curvas integrales, esto es, en todo el dominio del sistema de ecuaciones diferenciales, sin intentar conservar la escala. Consideremos el espacio en el que un sistema dado define un campo de direcciones como el espacio de la fase de un proceso físico. Entonces el esquema general de las curvas integrales correspondientes al sistema de ecuaciones diferenciales dará idea del carácter de todos los procesos (movimientos) que pueden posiblemente ocurrir en el sistema. En las figuras 10-13 dimos una representación esquemática aproximada del comportamiento de las curvas integrales en el entorno de un punto singular aislado.

Uno de los problemas fundamentales de la teoría de las ecuaciones diferenciales es el de encontrar un método tan sencillo como sea posible para construir un esquema del comportamiento de la familia de curvas integrales de un sistema de ecuaciones diferenciales dado en todo el dominio de definición, con el fin de estudiar el comportamiento global de las curvas integrales del sistema de ecuaciones diferenciales. Este problema apenas ha sido tocado para espacios de dimensión superior a 2, y no está todavía resuelto ni para ecuaciones de la forma

$$\frac{dy}{dx} = \frac{M(x, y)}{N(x, y)}, \quad (64)$$

ni siquiera cuando $M(x, y)$ y $N(x, y)$ son polinomios.

⁴ A. M. Lyapunov, *The general problem of stability of motion*. («El problema general de la estabilidad del movimiento.»)

De ahora en adelante, se supondrá que las funciones $M(x, y)$ y $N(x, y)$ tienen derivadas parciales continuas de primer orden.

Si todos los puntos de un dominio simplemente conexo G , en el que está definido el segundo miembro de la ecuación diferencial (64), son ordinarios, la familia de curvas integrales puede representarse esquemáticamente como una familia de segmentos rectilíneos paralelos, ya que en este caso sólo pasará una curva integral por cada punto, y dos curvas integrales no pueden cortarse. Para una ecuación de tipo más general, que puede tener puntos singulares, la estructura de las curvas integrales puede ser mucho más complicada. El caso en el que la ecuación (64) tenga un conjunto infinito de puntos singulares (es decir, puntos donde se anulen simultáneamente el numerador y el denominador) puede excluirse, al menos cuando $M(x, y)$ y $N(x, y)$ son polinomios. Así, pues, nos limitaremos a considerar los casos en los que la ecuación (64) tenga un número finito de puntos singulares aislados. El comportamiento de las curvas integrales que se hallan cerca de uno de estos puntos singulares constituye el elemento esencial en el que se basa una representación esquemática del comportamiento de todas las curvas integrales de la ecuación.

Un elemento muy típico en el esquema para el comportamiento de todas las curvas integrales de la ecuación (64) está formado por los llamados *ciclos límite*. Consideremos la ecuación

$$\frac{d\varrho}{d\phi} = \varrho - 1, \quad (65)$$

donde ϱ y ϕ son las coordenadas polares en el plano Oxy .

La colección de todas las curvas integrales de la ecuación (65) viene dada por la fórmula

$$\varrho = 1 + Ce^{\phi}, \quad (66)$$

donde C es una constante arbitraria, distinta para diferentes curvas integrales. Para que ϱ sea no negativo, es necesario que ϕ tome valores no mayores que $-\ln |C|$, $C < 0$. La familia de curvas integrales consistirá en:

1. La circunferencia $\varrho = 1$ ($C = 0$);
2. Las espirales que parten del origen y que tienden a esta circunferencia desde el interior cuando $\phi \rightarrow -\infty$ ($C < 0$).
3. Las espirales que se aproximan a la circunferencia $\varrho = 1$ desde el exterior cuando $\phi \rightarrow -\infty$ ($C > 0$) (Fig. 14).

La circunferencia $\varrho = 1$ se llama ciclo límite de la ecuación (65). En general, una curva integral cerrada l se llama *ciclo límite* si puede ser encerrada en un disco tal que todos sus puntos sean ordinarios para la ecuación (64) y totalmente ocupado por curvas integrales no cerradas.

En la ecuación (65) puede verse que todos los puntos de la circunferencia son ordinarios. Esto significa que una pequeña parte de un ciclo límite no se diferencia de una pequeña parte de cualquier otra curva integral.

Toda curva integral cerrada en el plano Oxy da una solución periódica $[x(t), y(t)]$ del sistema

$$\frac{dx}{dt} = N(x, y) \quad (67)$$

$$\frac{dy}{dt} = M(x, y),$$

que describe la ley de variación de un sistema físico. Las curvas integrales del plano de las fases que se aproximan a un ciclo límite cuando $t \rightarrow +\infty$

son movimientos que para $t \rightarrow +\infty$ se aproximan a movimientos periódicos.

Supongamos que para todo punto (x_0, y_0) suficientemente próximo a un ciclo límite l se tiene la situación siguiente: si (x_0, y_0) se toma como punto inicial (es decir, para $t = t_0$) de la solución del sistema (67), entonces la curva integral correspondiente descrita por el punto $[x(t), y(t)]$ se aproxima, cuando $t \rightarrow +\infty$, al ciclo límite l en el plano Oxy . (Esto significa que el movimiento en cuestión es aproximadamente periódico.) En este caso el correspondiente ciclo límite se dice *estable*. Las oscilaciones que actúan de este modo con respecto a un ciclo límite corresponden físicamente a las auto-oscilaciones. En algunos sistemas auto-oscilantes pueden existir varios procesos oscilatorios estables de diferentes amplitudes, algunos de

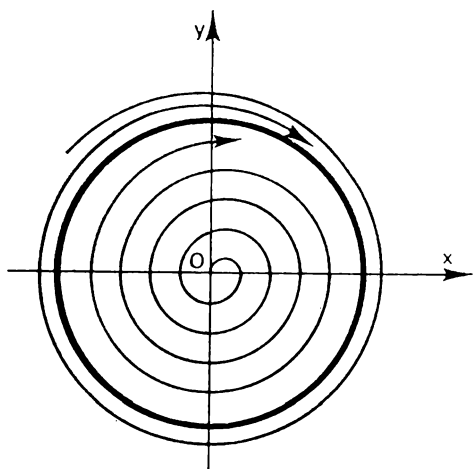


FIG. 14

los cuales vendrán establecidos por las condiciones iniciales. En el plano de las fases de tales «sistemas auto-oscilantes» existirán los correspondientes ciclos límites si los procesos que tienen lugar en estos sistemas vienen descritos por una ecuación de la forma (67).

El problema de encontrar (aunque sea aproximadamente) los ciclos límites de una ecuación diferencial dada no ha sido todavía resuelto de un modo satisfactorio. El método más utilizado para resolver este problema es el que sugirió Poincaré: construir «ciclos sin contacto». Se basa en el siguiente teorema. Supongamos que se pueden encontrar dos curvas cerradas L_1 y L_2 (ciclos) en el plano Oxy con las propiedades siguientes:

1. La curva L_2 se encuentra en la región limitada por L_1 .
2. En el anillo Ω , comprendido entre L_1 y L_2 , no hay puntos singulares de la ecuación (64).
3. L_1 y L_2 tienen tangentes en todo punto, y las direcciones de estas tangentes no coinciden nunca con las del campo de direcciones de la ecuación dada (64).
4. Para ninguno de los puntos de L_1 y L_2 cambia de signo el coseno del ángulo entre las normales interiores al contorno del dominio Ω y el vector de componentes $[N(x, y), M(x, y)]$.

Entonces entre L_1 y L_2 hay al menos un ciclo límite de la ecuación (64).

Poincaré llamó a las curvas L_1 y L_2 *ciclos sin contacto*.

La demostración de este teorema se basa en el siguiente hecho, bastante evidente. Supongamos que para t decreciente (o para t creciente) todas las curvas integrales

$$x = x(t), \quad y = y(t)$$

de la ecuación (64) (o, lo que es lo mismo, de las ecuaciones (67), donde t es un parámetro), que cortan a L_1 o L_2 , penetran en el anillo Ω comprendido entre L_1 y L_2 . Las curvas integrales deben entonces tender necesariamente a alguna curva cerrada l situada entre L_1 y L_2 , puesto que ninguna de las curvas integrales situadas en el anillo puede abandonarlo, y no hay puntos singulares en él.

Pero el problema de encontrar ciclos sin contacto es también complicado y no se conocen métodos generales para resolverlo. En casos particulares se han podido encontrar ciclos sin contacto, probándose así la existencia de ciclos límite.

En radiotécnica es importante encontrar ciclos límite (procesos auto-oscilantes) para la ecuación (16) del generador electrónico.

Para las ecuaciones del tipo (16), N. M. Krylov y N. N. Bogoljubov encontraron un método, hace ya veinte años, para el cálculo aproximado de los ciclos límites de esta ecuación. Aproximadamente al mismo tiempo los físicos soviéticos L. Y. Mandelstam, N. D. Papaleksi y A. A. Andronov dieron una demostración de la posibilidad de aplicar el llamado método del pequeño parámetro que, con cierta extensión, había sido utilizado anteriormente, aunque sin justificación rigurosa. Andronov fue también el primero en hacer un uso práctico y sistemático, en el análisis de los sistemas auto-oscilantes, de los métodos teóricos desarrollados ya por Lyapunov y Poincaré. De este modo obtuvo resultados importantes.

Como ya se indicó, en física tienen gran importancia los sistemas «poco sensibles» (ver §3). Andronov, junto con L. S. Pontryagin, elaboró un catálogo de los elementos a partir de los cuales se podía construir un mapa completo del comportamiento de las curvas integrales en el plano Oxy de una ecuación diferencial «poco sensible» de la forma (64). Desde hacía tiempo se sabía, por ejemplo, que un centro cercano a un punto singular se destruye fácilmente con pequeños cambios en las ecuaciones (64). Así, en la construcción de un mapa del comportamiento de las curvas integrales de la ecuación (64), no puede haber un centro, es decir, una familia de curvas integrales cerradas que rodee a un punto singular, si la ecuación es «poco sensible».

La cuestión del comportamiento global de las curvas integrales está todavía lejos de su solución final. La cuestión análoga, y probablemente más sencilla, de la forma de las curvas algebraicas reales en el plano, es decir, las curvas definidas por la ecuación

$$P(x, y) = 0,$$

donde $P(x, y)$ es un polinomio de grado n , está también lejos de una solución completa. La forma de estas curvas sólo se conoce para $n < 6$.

Las soluciones del sistema (64) definen movimientos en el plano. Si se reemplaza cada punto (x_0, y_0) del plano por el correspondiente punto $[x(t, x_0, y_0), y(t, x_0, y_0)]$, donde $x(t, x_0, y_0)$ e $y(t, x_0, y_0)$ son la solución del sistema (64) con las condiciones iniciales $x = x_0$ e $y = y_0$ para $t = t_0$, se obtiene una transformación de los puntos del plano dependiente del parámetro t . Transformaciones similares, dependientes de un parámetro, junto con los movimientos que generan, pueden estudiarse sobre una esfera, un toro u otras variedades. Las propiedades de estos movimientos se estudian en la teoría de los

sistemas dinámicos. En un entorno de todo punto estos movimientos son las soluciones de un cierto sistema de ecuaciones diferenciales. En la pasada década se ha desarrollado la teoría de los sistemas dinámicos en base a los trabajos de V. V. Stepanov, A. Ja. Hincin, N. N. Bogoljubov, N. M. Krylov, A. A. Markov, V. V. Nemyckii y otros, y también en los de G. D. Birkhoff y otros matemáticos.

En este capítulo se ha dado una breve descripción del estado actual de la teoría de las ecuaciones diferenciales ordinarias y se han intentado describir los problemas que se estudian en esta teoría. Dicha descripción en modo alguno pretende ser completa. Se han tenido que omitir consideraciones sobre muchas ramas de la teoría que aparecen al estudiar problemas más especiales o que requieren un conocimiento matemático más amplio del que se supone posee el lector. Por ejemplo, no se ha tocado la importante y amplia rama que trata de la teoría de las ecuaciones diferenciales de variable compleja. No ha habido oportunidad de estudiar la teoría de los problemas de contorno y, en particular, la de las autofunciones, de gran importancia en las aplicaciones.

Tampoco se ha podido prestar mucha atención a los métodos aproximados para la solución numérica o analítica de las ecuaciones diferenciales. Para estas cuestiones se recomienda al lector la consulta de literatura especializada.

INDICE ANALITICO

A

- Abel, N. H., I: 99, 318, 332, 333, 335; II: 171; III: 332, 343
- Abscisa, I: 260
- Adams, I., I: 21
- Adición, III: 26
- Aditiva, propiedad, II: 238
- Aditividad total de la medida, III: 43
- Ado, I. D., III: 396
- Afín, clasificación de curvas, figuras, superficies, etc., I: 283
- espacio, III: 77, 164
- propiedades de las figuras, I: 283
- Agregado; véase Conjunto
- Al-Biruni, I: 60
- Aleksandrov, A. D., I: 302; II: 140
- Aleksandrov, P. S., III: 41, 195, n. 3 p. 239, 260, 261, 265, 266, 373
- Alexander, J., III: 260, 261
- Algebra, I: 63-4, 82-3, 315-6; III: 309, 385
- abstracta, I: 319
- asociativa, III: 385
- axiomática, I: 319
- base de, III: 387
- con división, III: 394
- estructura de, III: 393
- topológica, III: 261
- ideal de, III: 392
- isomorfa, III: 388
- de Lie, III: 395
- lineal, III: 55
- de matrices, III: 390-391
- nihilpotente, III: 394
- rango de, III: 387
- semisimple, III: 394, 398
- simple, III: 394
- sistemas hipercomplejos, III: 310, 385
- teorema fundamental del, I: 338-52
- vectorial, I: 261; III: 381
- Algebraico(a)
- complemento, III: 87
- entero, II: 240
- operación, III: 406-407
- solución, I: 320
- variedad, III: 402
- Algebrización, I: 316

- Algoritmo, II: 238, 239
 Euclídeo, II: 238
 Al-Kharizmi, I: 61, 62, 63, 317
 Alternado, grupo, III: 327
 Análisis, I: 68, 73, 83, 91
 funcional, I: 84, 311-12; III: 265, 267
 Analítica, geometría, I: 68, 70, 225, 228-9
 en el espacio, I: 266-7
 infinito dimensional, I: 311
 prolongación, II: 222, 224, 225
 superficie, II: 125
 teoría de números, II: 240
 Analiticidad de una función compleja, II: 219
 Andronov, A. A., I: 396, 424; III: 254, 255
 Angulo, entre vectores, III: 77, 270
 de paralelismo, III: 133-134, 136, 138
 Anillo, III: 398
 de clases residuales, III: 400
 factor, III: 400
 no conmutativo, III: 403
 de operadores, III: 404
 principales, III: 401
 semisimple, III: 404
 sin divisores de cero, III: 401
 Apolonio, I: 57, 58, 70
 Aproximación, por defecto, I: 364, 366
 por exceso, I: 364, 367
 fórmula de, II: 356
 de funciones, II: 311
 de integrales definidas, II: 324
 métodos de, II: 355
 óptima, II: 313, 337
 óptima uniforme, II: 313, 330, 340
 por polinomios, II: 314, 333
 Arco, longitud de, II: 82
 Argumentos de un número complejo, I: 340
 Aristóteles, I: 92
 Aritmética, I: 24
 Armónica, función, II: 30
 Arquímedes, I: 33, 57, 58, 59, 71, 91, 97, 167; III: 31
 principio de, III: 31
 Aryabhata, I: n. 18 p. 60
 Asintótica, convergencia, III: 135
 Asintótica, ley, II: 246
 Asintótica en un punto, II: 77
 Asociados, elementos, III: 401
 Asociativa, álgebra, III: 385
 Asociatividad, III: 26
 Autofunción, II: 47; III: 295
 Autointersección, III: 238
 Automorfismo, III: 317
 Automorfismo, grupo de, III: 345
 Autooscilaciones u oscilaciones propias, I: 391
 Autovalor(es)
 espectro continuo de, III: 304
 de operadores, III: 301
 de una transformación lineal, III: 105
 Autovector(es), III: 104-105
 de operadores, III: 301-302
 de una transformación lineal, III: 104-105
 Autovibraciones, I: 391; III: 293
 Axioma(s), de adición, III: 26
 de continuidad, I: 347; III: 29, 153
 de distributividad, III: 27
 de un espacio lineal, III: 68
 de geometría, III: 151
 de incidencia, III: 152
 del movimiento, III: 153
 de multiplicación, III: 26
 de orden, III: 27, 152
 de paralelismo, III: 153
 de teoría de la probabilidad, II: 272
 Axiomática, álgebra, I: 319
- B**
- Banach, S., I: 84
 Base, de un álgebra, III: 387
 de un espacio, III: 71-72
 Básico, intervalo, II: 257
 Bernoulli, D., I: 74, 98, 369; III: 281
 Bernstein, S. N., II: 290, 307, 314, 329, 336, 340
 Bernstein, polinomio de, II: 339, 340
 Bertrand, J. L. F., II: 245

Bessel, F. W., I: 222
 Bézout, E., I: 318
 Bhaskara, I: n. 18 p. 60
 Birkhoff, G., I: 425; III: 260
 Blaschke, W., II: 143
 Bogoljubov, N. N., I: 424, 425; II: 302
 Boltjanski, V. G., III: n. 12 p. 260, 261
 Bolyai, F., III: 127, 131, 375
 Bolyai, J., I: 80; III: 131
 Bolzano, B., I: 83, 347, 348; III: 16, 36
 Bonnet, P. O., II: 128
 Borel, E., I: 83; II: 314
 Boyle, K., I: 105, 293; II: 29; III: 174
 Brahe, T., I: 242
 Brahmagupta, I: n. 18 p. 60
 Braquistócrona, II: 145, 146
 Briggs, H., I: 64
 Brouwer, L. E. J., III: 259, 260
 Budan, I: 357
 Bunjakovski, V. Ja., III: 79, 80, 276, 277

C

Cadena, III: 248
 bidimensional, III: 249
 unidimensional, III: 248
 Caja canónica, III: 108
 Cálculo, diferencial, I: 71, 166
 integral, I: 71, 72, 166
 matricial, III: 117
 técnicas de II: 355
 Cálculo de variaciones, II: 145
 funciones admisibles, II: 152
 problema variacional en, II: 157, 159
 soluciones aproximadas en, II: 164, 165
 Calor, vector conducción de, II: 21
 Cambio de variable, I: 175-6
 Campo(s)
 algebraico, III: 27
 continuo de direcciones, III: 252
 de descomposición, III: 344
 de direcciones, I: 398; III: 252-254
 escalar, II: 16

vectorial, II: 15; III: 252
 continuo, III: 252
 Canónica, forma de una ecuación, I: 254
 de una matriz, III: 111-112
 forma cuadrática, III: 113
 matriz de Jordan, III: 108
 Cantor, G., I: n. 14 p. 47, 79, 83; III: 18, 29, 31-34, 37, 41, 259
 conjunto perfecto de, III: 39
 principio de continuidad de, III: 29
 Cantoriana, continuidad, III: 29
 de una figura, III: 194
 Caracteres, de grupos, III: 367
 Característica(o), ecuación, de una ecuación lineal homogénea, I: 389
 ecuación cuadrática, I: 292
 frecuencia; véase Autofunciones
 función; véase Autofunciones
 oscilaciones, I: 391
 Cardano, G., I: 318, 321, 322, 324, 325
 Cartan, E., III: 214, 358, 398
 Cauchy, A., I: 74, 83, 99, 117, 119, 171, 172; II: 180, 183, 187, 194, 198, 204; III: 16, 30, 31, 79, 80, 276, 277
 principio de, III: 30
 Cauchy-Riemann, ecuaciones de, II: 179, 206
 Cavalieri, B., I: 71
 Cayley, A., III: 169, 351
 Centro, I: 415
 Cero, I: 32
 Cerrado(a), clase de curvas o superficies, II: 139
 conjunto, III: 37, 262
 sistema ortogonal de funciones, III: 287
 superficie, III: 236, 244, 246
 superficie convexa, II: 140
 Cíclica, permutación, III: 328
 Cíclico, subgrupo, III: 331
 Ciclo(s), fundamental, III: 251-252
 límite, I: 421; III: 253
 límite estable, I: 422
 longitud de, III: 329-330
 permutación en, III: 328
 en topología, III: 250
 uni-dimensional, III: 249

- Ciclotómicas, ecuaciones, I: 332
 Cilíndricas, funciones, III: 282
 Cilindro elíptico, I: 270
 hiperbólico, I: 270
 parabólico, I: 270
 Circulación de flujo, II: 192
 Circulatorio, movimiento, II: 191
 Círculo de convergencia, II: 173
 Círculo límite, III: 139
 Circunferencia, no orientada, III: 337
 orientada, III: 336
 Clairaut, A. C., II: 178
 Clase cerrada de curvas y superficies, II: 139
 Clase, III: 352
 a la derecha, III: 352
 a la izquierda, III: 352
 Cociente por la derecha, III: 332
 Cociente por la izquierda, III: 332
 Cociente de diferencias de una función, II: 60
 Codazzi, II: 80, 128, 134
 ecuaciones de, II: 80, 128, 134
 Coeficiente de dilatación, II: 196, 363
 Coeficiente de una matriz, III: 62
 de transformación, III: 101
 Cohn-Vossen, S. E., III: n. 2 p. 343
 Cohomología, III: 261
 Colón, I: 91
 Columna, de una matriz, III: 59
 Combinatoria, topología, III: 260, 261
 Complejo(a), raíz, I: 327
 variable(s), ecuaciones de Cauchy-Riemann para, II: 179
 coeficiente de dilatación en, II: 196
 coordenadas, I: 312
 espacio, I: 312
 número, III: 374-375, II: 167-183
 argumento del, I: 340
 Complejo, plano, I: 352
 Complejos, enteros, II: 265, 267
 Completo, sistema de sucesos, II: 273
 Compleitud de la recta numérica, III: 33
 composición de grupos, III: 347
 Computacional, matemática, I: 86
 Concavidad, I: 146-7
 Condición necesaria de convergencia, I: 209
 Conexión, III: 194
 Conexión de superficie, III: 235
 de variedades, III: 251
 Conforme, geometría, II: 143
 Conforme, transformación, II: 197, 200
 Congruencias, II: 137
 Congruencias de rectas, II: 137
 Congruentes, superficies, II: 128
 Cónicas, secciones, I: 252
 foco de, I: 248
 (Conjunto(s), abierto, II: 37
 acotado, III: 33-34
 inferiormente, III: 34
 superiormente, III: 34
 acotado, e infinito de puntos, III: 36
 adherente, III: 194, 263
 aplicación unívoca de, III: 353
 cardinalidad de un, III: 22
 cero-dimensional, III: n. 11 p. 260
 cerrado, III: 37
 colineal de puntos, I: 298
 complemento de un, III: 38
 definición inductiva de dimensión de, III: n. 11 p. 260
 densidad de un, III: 27
 envoltura de, III: 42
 extremo inferior de, III: 34
 extremo superior de, III: 30-34
 finito, III: 20
 de funciones lineales, III: 57
 igual, III: 19
 infinito, III: 20, 36
 intersección de, III: 19
 lineal, III: 33
 linealmente ordenado, III: 404
 medible, III: 43
 n-dimensional, III: n. 11 p. 260
 no acotado, III: 33
 numerable, III: 23-24
 números racionales, III: 24
 ordenado, III: 33, 404
 parcialmente ordenado, III: 404
 perfecto, III: 39
 teoría de, III: 17

- teoría descriptiva de, III: 40
- totalmente ordenado, III: 404
- unión de, III: 19
- vacío, III: 19
- Conmutación, III: 396
- Conmutador, III: 340
 - de elementos de grupo, III: 333
- Conmutadores, subgrupo de, III: 333
- Conmutatividad, III: 26
- Conmutativo, grupo, III: 332
- Cono de segundo orden, I: 271
- Constante de estructura, III: 388
- Continuo(a), aplicación
 - función no regular, II: 323
 - funcional, III: 298
 - funciones, I: 117
 - grupo, III: 356, 358
 - grupo de transformaciones, III: 356
 - en un intervalo, I: 118
 - lateralmente, I: 119
 - en un punto, I: 118
 - representación, III: 358
 - transformación, III: 195, 232, 233, 263; II: 126
 - de medio continuos, I: 284-285; III: 257
 - variable, I: 54
 - no numerabilidad del, III: 32
 - número cardinal del, III: 25
- Contorno, condiciones de, II: 31, 34
 - condiciones homogéneas de, II: 36
 - problema de, II: 29, 33, 155
- Contracción del plano a una línea, I: 276, 285
- Contracción, teoremas de, II: 364
- Convergencia, asintótica, III: 135
 - círculo de, II: 173
 - dominio de, II: 171
 - «en media», III: 283
 - rapidez de, II: 361
 - de series, I: 209
- Convergente, onda esférica, II: 43
- Convergentes, métodos, II: 361
- Convergentes, series, I: 209
 - absolutamente, I: 211
- Convexa hacia abajo, I: 147
- Convexidad, I: 146-147
- Convexo, cuerpo, II: 140-142
- Coordenadas, curvilíneas, II: 114, 129
 - método de, I: 229
 - polares, I: 310
 - de un punto, I: 227
 - rectangulares, I: 260
 - sistema de, I: 281
 - transformación de, I: 255; III: 102-103
 - vectoriales, I: 263; III: 74
- Copérnico, N., I: 91, 225
- Correlación, coeficiente de, II: 288
- Correspondencia, uno a uno, III: 21
- Corriente, línea de, II: 185
- Cortadura de Dedekind, III: 29
- Corte en funciones complejas, II: 182
- Cota, I: 260
- Cota superior, III: 30
- Coulomb, C. A., I: 76
- Criba de Eratóstenes, I: 58; II: 241
- Cuadrática, ecuación, I: 321
 - característica, I: 292
- Cuaterno, norma de, III: 380
- Cuaternos, parte vectorial de, III: 380, 382-383
 - parte escalar de, III: 380
- Cuaternos conjugados, III: 378-379, 380-381
- Cúbica, ecuación, I: 321
- Cubo tetradimensional, III: 177
- Cuerpo, III: 378
 - algebraico, III: 27
 - ordenado, III: 27
 - de base de álgebra, III: 344, 386
 - no simétrico, III: 319
 - simétrico, III: 318
- Curva(s), clases afines de, I: 283
 - clases cerradas de, II: 138
 - curvatura de, II: 87-90
 - ecuación de, I: 266
 - equidistante, III: 135
 - en el espacio, longitud del arco de, II: 83
 - familias de, II: 234
 - integral, I: 399; III, 253
 - irreducible, III: 403
 - longitud de, II: 83

matemática, II: 75
 n -dimensional, III: 184
 de nivel, I: 181, 267
 Curvatura, cero de una curva, II: 87, 90
 constante, III: 211
 del espacio, III: 209
 Gaussiana, II: 108
 media, II: 108
 principal, II: 103
 sobre una superficie, II: 101-108
 total; véase Curvatura Gaussiana
 Curvatura, tensor, III: 208
 Curvilínea(s), coordenadas, II: 114
 129
 integral, I: 201-203

CH

Chaplygin, S. A., II: 191, 193
 Chebotarev, H. G., I: 337, 338;
 III: 396, 402, 408
 Chebichev, P. L., I: 84, 87, 99,
 178; II: 79, 240, 243-246, 249,
 252, 263, 286-290, 329-334,
 335, 341, 350, 351; III: 282
 aproximación de, II: 313, 330-
 332, 340
 desigualdad de, II: 287
 polinomios de, II: 322, 323, 334
 Chech, E., II: 143
 Chudakov, N. G., II: 247

D

D'Alembert, J., I: 211; III: 127,
 281
 criterio de series convergentes,
 I: 211
 lema de, I: 349
 Dandelin, G. P., I: n. 14 p. 369
 Darboux, G., II: 80, 125
 Darwin, C. R., III: 131
 Da Vinci, L., I: 91; III: 160
 Dedekind, R., I: n. 14 p. 47, 83;
 II: 240; III: 29-32, 406
 Dedekind, continuidad de, I: 347;
 III: 29-30, 153
 cortadura, III: 28-29

 principio de, III: 29
 retículo, III: 406
 Definida, integral, I: 165
 Deformabilidad de una superficie,
 II: 142
 Deformación de una superficie, II:
 115, 125, 142
 Deformación uniforme de una
 membrana, II: 147
 del Ferro, Scipio, I: 321, 329
 de Jonquières, III: 245
 Demócrito, I: 40, 45, 53, 54
 Densidad lineal, III: 292
 Dependencia, funcional, II: 149
 lineal, III: 69
 Derivación, reglas de, I: 132
 término a término, I: 212
 Derivada, de un cociente, I: 134
 de una función, I: 124
 de orden superior, I: 145
 parcial, I: 182
 parcial cruzada, I: 184
 de un producto, I: 133
 segunda, I: 146
 de una suma, I: 132
 Desargues G., III: 160
 Desarrollo,
 de una función en serie trigono-
 métrica, II: 243; III: 284
 en un sistema ortogonal de fun-
 ciones, III: 279, 282
 Descartes, R., I: n. 13 p. 46, 59,
 63, 68, 69, 70, 85, 96, 226-231,
 233-234, 236, 237, 260, 316, 317,
 318, 355-358; II: 129; III: 243
 Descartes, ley de los signos de, I:
 355
 Descomposición, de números ente-
 ros, II: 265
 de una permutación en ciclos,
 III: 328
 simplicial, III: 247
 de una variedad, III: 403
 Desviación
 típica, II: 287
 de frecuencia, II: 286
 probable, II: 276
 Determinados, procesos, II: 299
 Determinante(s), de los coeficien-
 tes de una matriz, II: 89

- complemento algebraico en, III: 87
 - desarrollo de, III: 89
 - de una matriz, III: 83
 - menor, III: 87
 - de orden n , III: 86-87
 - Diagonal, proceso, III: 24
 - Diámetro, de una región, I: n. 34 p. 200
 - teoría de Newton de, I: 237
 - Diferencia de dos conjuntos, III: 20
 - Diferenciabilidad de una función, II: 177
 - Diferenciable, función, I: 129; II: 195
 - Diferencial(es), ecuación(es), I: 373 y sigs.
 - campo de direcciones, I: 398; III: 253-254
 - centro de, I: 415; III: 258
 - en derivadas parciales, I: 376
 - estable, I: 417
 - estocástica, II: 307-309
 - de Euler, II: 151-154
 - foco en, I: 415
 - lineal, I: 386-387
 - lineal homogénea, I: 387
 - orden de, I: 384
 - ordinaria, I: 373, 376, 383
 - «poco sensibles», I: 396
 - solución general de, I: 385
 - Diferencial de una función, I: 150
 - de varias variables, I: 185
 - Dilatación, coeficiente de, II: 196, 363
 - Dimensión(es), del espacio, III: 71
 - de una variedad, III: 242-243
 - Dimensión, teoría de, III: 259
 - Diofanto, I: 59, 60
 - Dirección, números de, I: 407
 - principal, II: 103
 - Directos, métodos, I: 402
 - Directriz de una parábola, I: 248
 - Dirichlet, P. G. L., I: 108; II: 74; III: 46, 49
 - función de, III: 46
 - Discontinua, función, I: 119
 - Discontinuidad débil, II: 68
 - Discreto, grupo de movimientos, III: 338, 342
 - Discreto, grupo simétrico, III: 338
 - Disjuntas, clases, en grupos, III: 349
 - Disjuntos, conjuntos, III: 20
 - Distancia, en un espacio n -dimensional, III: 262
 - entre funciones, III: 263-264
 - entre puntos, III: 33
 - Distribución de números primos, II: 242-244
 - Distributiva, ley, III: 20
 - Distributividad, III: 27
 - Divergencia, teorema de Gauss de; véase Ostogradski, fórmula de
 - Divergentes, series, I: 207
 - Divisibilidad, II: 236
 - División, álgebra con, III: 394
 - Divisor, de cero, III: 401
 - normal, III: 352
 - de uno, III: 401-402
 - Dominio, de convergencia, II: 171
 - de definición, II: 149; III: 47, 299
 - de una función, I: 103
 - integral, III: 401
 - de racionalidad, III: 344
 - simplemente conexo, II: 201
 - de valores, III: 191
 - de variación de variables, III: 191
 - Dualidad, teorema de, III: 261
 - Durero, A., III: 160
- E
- Ecuación(es), característica, homogénea lineal, I: 389
 - de Cauchy-Riemann para variables complejas, II: 179, 206
 - centro en, I: 415
 - ciclotómica, I: 333
 - de conservación de la energía calorífica, II: 22
 - de continuidad, II: 20
 - cuadrática, I: 321
 - de cuarto grado, I: 324
 - de una cuerda vibrante, II: 26
 - de una curva, I: 266
 - diferencial, I: 373-425

- construcción de soluciones, II: 43
 estocástica, II: 307-309
 estructuralmente estable, I: 417
 lineal, I: 386
 lineal homogénea, I: 387
 lineal no homogénea, I: 387, 392-395
 ordinaria, I: 373, 375, 384
 solución general de, I: 385
 forma canónica de, I: 255
 funcional, I: 374
 homogénea correspondiente, II: 43
 integral, III: 288-296
 homogénea, III: 292
 no homogénea, III: 294
 de segunda especie, II: 57
 de Laplace, II: 30, 194
 de una línea recta, I: 268
 del movimiento, II: 23
 de oscilación, II: 35
 parcial derivada, I: 376; II: 15
 soluciones admisibles o generalizadas de, II: 65, 69
 construcción de soluciones, II: 43
 paramétrica, II: 81
 del plano, I: 268
 de Poisson, II: 30
 de primer grado, I: 231, 320
 de Ricatti, I: 395
 secular, III: 106
 de segundo grado, I: 320
 solución algebraica de, I: 320
 de superficie, I: 266
 de tercer grado, I: 321
 Efimov, N. V., II: 127
 Egorov, D. F., III: 47
 Einstein, A., I: 301; III: 131, 200, 216, 223-227
 Eje, de grupo, III: 338-339
 de rotación, III: 312
 de simetría, III: 312
 Elemento(s), asociado, de un anillo, III: 401
 cero, III: 348
 conjugados, de un grupo, III: 349
 de un conjunto, III: 18
 de longitud, III: 205
 de una matriz, III: 58
 indescomponible, III: 401
 inverso, III: 99
 neutro, III: 348
 primo, III: 401
 de superficie, II: 120, 132
 unidad, III: 348-349
 Elipse, I: 240-241
 imaginaria, I: 254
 de inercia, I: 243-245
 Elipsoide, I: 269
 imaginario, I: 269
 Envoltente de una familia de curvas, II: 135
 Equivalentes, caminos, III: 360
 Eratóstenes, I: 58; II: 241
 criba de, I: 58; II: 241
 Escalar, I: 262; II: 16
 multiplicación, III: 80
 Escalar, producto, espacio n -dimensional, III: 271
 de vectores, I: 262; III: 78, 383
 Espacio, abstracto, III: 164
 transformación de, I: 280-281; III: 162
 base de, III: 71-72
 configuración de, III: 213
 de color, III: 185
 curvatura de, III: 209-210
 de curvatura cero, III: 211
 curvatura constante, III: 211
 dimensión de, III: 71, 259
 de dimensión infinita, III: 71, 191, 196, 273-279, 298
 Euclidiano, III: 80
 fásicos, I: 410; III: 187, 243
 funcional, III: 192, 265, 299
 de infinitas dimensiones, III: 299
 generalizado, III: 190
 de Hilbert, III: 196, 273-379
 lineal complejo, III: 76
 lineal general, III: 68
 lineal de dimensión infinita, III: 298
 lineal de dimensión finita, III: 386
 métrico, III: 196, 262-263
 multidimensionales, III: 168, 174
 n -dimensional, III: 175

- real, III: 215
 de Riemann, III: 199-202
 rotación de, III: 318
 solución, III: 97
 topológico, III: 194, 262, 265
 topológico abstracto, III: 164
 unitario, III: 82
 vectorial lineal, III: 68-69
 Espacio-tiempo, III: 222
 Espectro continuo, III: 304-305
 Esperanza matemática, II: 287
 Estimación de un error, II: 360
 Estabilidad de los procesos de
 cálculo, II: 369
 Euclides, I: 22, 34, 40, 41, 46,
 n. 13 p. 46, 57, 58, 63, 77; II:
 241, 244, 247; III: 123-130, 140,
 141, n. 7 p. 144, n. 10 p. 151,
 152, 154, 160, 164, 211, 215,
 n. 37 p. 215, 217, 220
 Euclides, postulado de las parale-
 las de, III: 124, 153
 Eudemo de Rodas, I: 39
 Eudoxio, I: n. 12 p. 46
 Euler, L., I: 75, 78, 98, 127, 201,
 253, 259, 260, 317, 318, 402, 403,
 407; II: 54, 78, 103, 105, 107,
 114, 125, 134, 151, 154-157,
 159, 161, 175, 176, 240, 242,
 244, 246, 249, 260, 315, 343,
 370; III: 243-247, 252, 256 281
 Euler, ángulos de, I: 269
 características de triangulación,
 III: 244
 ecuación diferencial de, II: 151,
 154
 fórmula(s) de, II: 175; III: 244
 función(es) de, II: 260
 método de las líneas quebradas
 de, I: 403
 teorema de, II: 103, 105
 Euler-Poincaré, fórmula de, III:
 252
 Excentricidad de la sección cónica,
 I: 251
 Existencia de una solución, I: 401
 Extremas, condiciones, II: 150
 Extremos, valores de funcionales,
 II: 149
 Extrínseca, propiedad de una su-
 perficie, II: 127
- F
- Faber, G., II: 323
 Factor irreducible, I: 334
 Factorización de un polinomio, I:
 326
 Falsa, posición, I: 366, 368
 Familia de curvas, II: 134
 n -paramétrica, II: 134
 de superficies, II: 134
 Fedorov, E. S., I: 42, 82, 87; III:
 n. 23 p. 180, 311, 340, 342, n. 2
 p. 343
 Fejér, L., II: 349, 352
 Fejér, suma de, II: 349
 Fermat, P., I: 226, 316, 317; II:
 265; III: 402
 Ferrari, I: 63, 324
 Figuras, clasificación afín de, I:
 283
 planas infinitas, III: 337
 propiedades afines de, I: 283
 Finikov, S. P., II: 143
 Finsler, R., I: 81; III: 214
 Flujo, II: 20
 laminar, II: 188
 rotacional, II: 26
 vector de, II: 20
 Foco, en ecuaciones diferenciales,
 I: 415; III: 258
 de una sección cónica, I: 247
 Fok, V. A., III: 226
 Fokker, II: 307
 Forma cuadrática, III: 109, 114
 canónica, III: 113
 equivalente, III: 111
 fundamental, III: 204
 inercia de una, III: 113
 primera fundamental, II: 131-134
 segunda fundamental, II: 133
 Forma lineal, III: 58
 Fórmula, de los incrementos fini-
 tos, I: 154
 fundamental del cálculo integral,
 I: 205
 Fourier, J. B. J., I: 358; II: 44,
 54, 58, 315, 339, 341, 344, 347-
 349; III: 53, 196, 304
 Fourier, coeficientes de, II: 344
 integral de, III: 304
 series de, II: 341-349

- Fracciones, origen de, I: 43
 Frecuencia característica, *véase* Autofunción
 Frecuencias, espectro de, III: 293
 Frecuencia propia, III: 293
 fundamental, II: 51
 Frenet, J. F., II: 79
 Fridman, A. A., III: 227
 Frontera, de una cadena, III: 249
 de un conjunto, III: 194; *véase también* Punto frontera
 de un triángulo orientado, III: 247-248
 Fubini, G., II: 143
 Función(es), admisible, II: 152
 analítica, I: 219; II: 170
 completa, II: 224
 en intervalos, I: 220
 en un punto, II: 177
 representación de, II: 221
 aproximación de, II: 311
 aproximadora, II: 331
 armónica, II: 30
 máximo o mínimo absoluto de, II: 32
 característica, II: 47; III: 49
 de flujo, II: 187
 cociente de diferencia de, II: 60
 compleja, II: 167-231
 compuesta, II: 130
 constante a trozos, II: 68
 continua, I: 117
 a un lado, I: 119
 no regular, II: 139, 323
 continuamente diferenciable, II: 71; III: 55; *véase también* función continua
 de corriente, II: 185
 de cuadrado integrable, III: 299
 definición implícita de una, I: 181
 derivada de, I: 121
 derivable, I: 129; II: 195
 diferencial de, I: 150
 de Dirichlet, III: 46
 discontinua, I: 119
 dominio de, I: 103
 elemental, I: 139; II: 169
 entera, II: 216
 esférica, III: 281
 de Euler, II: 260
 exponencial, II: 174
 de matrices, III: 118
 fraccionaria, II: 218
 de una función, I: 137
 gráfica de, I: 103, 140-9
 de Green, III: 288-292
 impar, II: 237
 implícita, definición de, I: 181
 implícita, derivación de, I: 187
 integrable, I: 172, 196
 integral, II: 215
 inversa, I: 135
 de Lagrange, II: 70
 límite, III: 16
 lineal, I: 102; III: 55
 lineal homogénea, III: 56
 lisa, II: 68; III: 55, 98
 lisa a trozos, II: 68
 de una matriz, III: 117
 argumento, III: 117
 medible, III: 45
 multiforme, II: 181, 225
 no regular, II: 139
 ortogonales, III: 278
 ortogonales dos a dos, III: 279
 par, II: 237
 prolongación de, en un dominio complejo, II: 170
 regular, II: 322
 sistema ortogonal cerrado de, III: 287
 sistema ortogonal infinito de, III: 287
 de teoría de números, II: 260
 teoría trascendente de, II: 195
 trascendente, II: 217
 trigonométrica, II: 174
 de una variable compleja, II: 174
 de variable real, III: 15
 de varias variables, I: 179
 vectorial, II: 130
 Funcional, I: 155; III: 297

G

- Galerkin, B. G., II: 58, 59, 73, 358
 Galerkin, método de, II: 58
 Galileo, I: 65, 68, 70, 92, 93, 225, 301; II: 300, 301

- Galois, E., I: 82, 318, 330-8; III: 343-347
- Galois, grupo de, I: 355; III: 343-345
- teoría de, I: 332-6
- Gauss, K. F., I: 98, 318, 332, 336; II: 111-114, 128, 130, 134, 240, 244-330; III: n. 2 p. 126, 127, 131, 199, 209, 210, 216
- curvatura de, II: 108, 111; III: 209-210
- teorema de la divergencia de; véase Ostrogradski, fórmula de
- Gelfand, I. M., III: 358
- Gelfond, A. O., II: 240
- Generadores, de grupos, III: 363
- libres, III: 364
- de subgrupo, III: 330
- Generalizadas, soluciones de ecuaciones diferenciales parciales, II: 67, 69
- Generadores, de grupo, III: 363
- Generatrices rectilíneas, I: 274
- Género de una superficie, II: 230; III: 236
- Geodésica, II: 122; III: 143, 206
- Geodésica(o), curvatura, II: 119
- línea, II: 121
- local, II: 137
- superficie, III: 210
- segmento, II: 119
- Geometría, I: 38, 80
- abstracta, III: 215
- afín, III: 162
- algebraica, I: 311
- analítica, I: 68, 70, 225, 228, 311
- en el espacio, I: 266-7
- analítica de infinitas dimensiones, I: 311
- axiomas de la, III: 151
- conforme, II: 143
- diferencial, II: 77, 80-82, 143
- afín, II: 143
- clásica, II: 137
- proyectiva, II: 143
- euclidiana, III: 141
- global, II: 137
- intrínseca, II: 114-134; III: 201
- de Lobachevski, III: 132
- n -dimensional, III: 169
- no euclidiana, III: 123
- proyectiva, I: 293, 299; III: 159
- riemanniana, III: 199
- Geométrico(a), figura, I: 38
- interpretación, I: 398-400
- Gibbs, J. W., III: 174
- Gjunter, N. M., II: 72
- «Global», I: 420; II: 77, 136
- Goldbach, C., II: 249, 256, 259, 264
- Goldbach, conjetura de, II: 249, 256
- Gradiente de temperatura, II: 28
- Grado, de una representación mediante matrices, III: 367-392
- de simetría, III: 319
- Graduada, representación, III: 370
- Graeffe, I: n. 14 p. 369
- Graeffe, método de, I: 369
- Gráfica, de una función, I: 103, 104, 140-149
- de un polinomio, I: 364
- Grassmann, H., II: 129; III: 169
- Green, G., III: 288
- Green, función de, III: 288-292
- teorema de; véase también Ostrogradski, fórmula de
- Grupo(s),
- abeliano, III: 332
- abstracto, III: 373
- alternado, III: 327
- de automorfismos, III: 345
- axiomas de, III: 348
- de caminos, III: 361
- carácter de, III: 367
- clases disjuntas en, III: 349
- composición de, III: 347
- conmutativo, III: 332
- caracteres de, III: 261
- continuo, III: 356, 358
- cristalográfico, III: 334, 340, 342
- definición, general, III: 347
- derivado, III: 333
- de ecuaciones algebraicas, III: 344
- eje de, III: 338-339
- elemento conjugado en, III: 349
- factor, III: 351
- de Fedorov, III: 334, 340, 342
- finito, III: 373
- finito simple, III: 358
- fundamental, III: 359

- de Galois, I: 355; III: 343-344
 generadores de, III: 363
 infinito, III: 373
 isomorfo, III: 349
 libre, III: 364
 de Lie, III: 356, 358
 local, III: 396
 de movimientos, III: 338
 multiplicación de, III: 347
 de nudos, III: 365, 366
 operación de, III: 348
 orden de, III: 325
 de permutaciones, III: 327
 plano de Fedorov, III: 334, 340, 341, 342
 relaciones de definición de, III: 363
 representación de, III: 367
 representación graduada en, III: 370
 representación por matrices, III: 367
 de rotaciones, III: 326
 simétrico, de permutaciones, III: 327, 347
 suma de, III: 348
 simple, III: 352
 topológico, III: 358
 de transformaciones, III: 321-334, 347
 en geometría diferencial, II: 143
 Gyaseddin Jamschid, I: n. 18 p. 60, 62
- H
- Hadamard, J., II: 246
 Hamilton, W. R., II: 70-72, 130, 160
 Hamilton, principio de; véase Principio de Ostrogradski-Hamilton
 Hardy, G. H., II: 240
 Hermitiano, espacio vectorial; véase Espacio unitario
 Hertz, H., I: 21
 Hilbert, D., I: 84, 99, 337, 338
 espacio de, III: 196, 273-279
 Hincin, A. Ja., I: 425; II: 289
 Hiparco, I: 57
 Hipérbola, I: 247
 rectangular, I: 253
 Hiperboloide, I: 269, 273
 de revolución, I: 273
 Hipercomplejos, números, III: 374-385
 sistemas, III: 310, 385
 Hiperplano, III: 171
 tangente, III: 184
 Hipócrates, I: 56
 Holónomo, sistema, II: 160
 Homeomorfas, figuras, III: 235
 Homología, III: 250
 Homológica, independencia de los ciclos, III: 250
 teoría¹ de la dimensión, n. 11 p. 260
 Homólogo a cero, III: 250
 Homomorfa, aplicación, III: 353
 imagen, III: 256
 Homomorfismo, III: 353
 Homotópica, topología, III: 261
 Hooke, R., III: 290
 Hopf, H., III: 256, 260, 261
 Hopf, teorema de, III: 256
- I
- Ideal, III: 399
 bilátero, III: 393
 cero, III: 393
 de elementos de un álgebra III: 392
 a derechas, III: 392
 a izquierdas, III: 393
 minimal, III: 393
 principal, III: 399
 indescomponible, III: 403
 Ideales, líneas, I: 297
 puntos, I: 298
 Imagen especular, III: 311
 Inconmensurables, magnitudes, I: 44
 Incremento, I: 118, 150
 Independencia, III: 57
 lineal, III: 69
 Independientes, sucesos, II: 275
 Índice, de un punto singular, III: 256

de un punto singular aislado, III: 256
 Infinitamente grande, I: 112
 Infinitésimo, I: 110
 de orden superior, I: 151
 Inflexión, punto de, I: 148
 Inicial, condición, II: 34
 estado, I: 402; II: 29
 valor de un problema, I: 402; II: 29
 Integración, de caminos, II: 210
 por partes, I: 178
 Integral, I: 163
 de acción, II: 161
 de Cauchy, II: 211
 a lo largo de una curva, II: 208
 curvilínea, I: 201-3
 definida, I: 165
 doble, I: 200
 indefinida, I: 173; II: 210
 de Lebesgue, III: 47-53
 de línea, I: 201; II: 207
 sobre un contorno cerrado, II: 211
 múltiple, I: 197
 de Riemann, I: 165; III: 47
 de superficie, I: 201
 Interpolación, lineal, I: 366, 368
 polinomios de, II: 316, 335
 Intersección, de conjuntos, III: 19-20
 de sucesos, II: 274
 Intervalo, III: 33
 abierto, III: 33
 adyacente, III: 38
 básico, II: 259
 cerrado, III: 33
 encajado, III: 29
 impropio, III: 33
 Invariante, I: 288
 topológico, III: 362
 Inverso, de un cuaternio, III: 381
 de una matriz, III: 93
 Irreducibilidad de un polinomio, I: 334
 Irrotacional, movimiento en un fluido, II: 187
 Isomorfa(s), aplicación, III: 350
 Isomorfismo, III: 349
 local, III: 396
 Iteración, método de, II: 362

J

Jackson, D., II: 314, 340
 Jordan, C., III: 108, 121
 Jukovski, N. E., I: 22

K

Keldys, L. V., III: 41
 Keldysch, M. V., II: 340; III: 305
 Kepler, J., I: 70, 71, 91, 96, 226, 242, 243, 375
 Killing, III: 358, 398
 Kirchhoff, G. R., II: 389
 Kiselev, A. P., I: 40
 Klein, F., I: 99, 338; III: 144, 150-154, 163, 200, 214, 225, 238, 257, 258, 358
 Klein, botella de, III: 256-257
 Kolmogorov, A. N., II: 307, 314, 315, 348; III: n. 1 p. 25, 41, 261
 Kolosov, G. V., II: 195
 Korkin, A. N., II: 314
 Krasnoselski, M. A., III: 265
 Krylov, A. N., I: 372, 405
 Krylov, N. M., I: 424, 425
 Kuiper, N. H., II: 142; III: n. 8 p. 144
 Kummer, E. E., II: 240; III: 402
 Kurnakov, N. S., III: 174
 Kurosch, A. G., I: 320; III: 373

L

Lagrange, J. L., I: 74, 78, 154, 193, 194, 195, 196, 260, 261, 318, 328-32, 334, 335, 337; II: 70, 240, 320, 329; III: 16, 127, 169, 223, 343, 352
 Lagrange, fórmula de, I: 154
 fórmula de interpolación de, II: 320
 función de, II: 70
 multiplicadores de, I: 78, 193
 resolventes de, I: 330
 teorema de, III: 352
 Laguerre, E., I: 260
 Lambert, J. A., III: 125, 126

- Laminar, flujo, II: 188
 movimiento, II: 183
 Laplace, P. S., II: 30-32, 40, 43, 52, 54, 62, 68, 194, 284, 303, 359, 382, 383
 ecuación de, II: 30
 Laplaciana, II: 30
 Lappo-Danilevski, I. A., III: 121
 Laptev, G. F., II: 143
 Lavrentiev, M. A., I: 404; II: 207, 314, 340; III: 41
 Lebesgue, H. I.: 83, n. 25 p. 169, 197; II: 72, 314, 348, 353; III: 17 41, 47-53, 279, 287
 suma integral de, III: 52
 Lefschetz, S., III: 260
 Legendre, A. M., II: 244, 330; III: 125, 126, 282
 polinomios de, II: 330; III: 282
 Leibnitz, G. W., I: 68, 71, 96, 97, 127, 167, 169, 170, 171, n. 27 p. 171, 173, 316; II: 78, 135; III: 131
 Leverrier, U. J. J., I: 21, 375
 Ley(es) de inercia, III: 113
 de los grandes números, II: 280, 289
 de probabilidad, II: 269
 triangular, III: 264
 L'Hôpital, II: 78
 Lie, S., I: 320; III: 261, 356-358, 389, 395-398
 Lie, álgebra de, III: 395
 grupo de, III: 356-358
 representación continua de, III: 358
 Límite, I: 72, 108-17
 de una sucesión de matrices, III: 117
 uniforme, II: 69
 Línea, ecuaciones de, I: 268
 infinitamente distante, III: 239
 en el infinito, III: 239
 impropia, III: 239
 integral, I: 399
 Linnik, Ju. V., II: 247; III: 385
 Lisma, II: 142
 Liouville, J., I: 333, 395
 Littlewood, J. E., II: 240
 Liapunov, A. M., I: 74, 84, 99, 416-9, 424; II: 58, 290; III: 41
 estable en el sentido de, I: 417
 no estable en el sentido de, I: 417
 Ljusternik, L. A., II: 138; III: 260, 262, 265
 Lobachevski, N. I., I: 22, 40, 80, 81, 99, 108, 310, 318, 369, 371, 372; III: 123, 127-159, 164, n. 28 p. 193, 211, 215, 216, 217, 220, 221, 222, 227
 geometría de, III: 132
 curva equidistante en la, III: 135
 método de, I: 369
 postulado de las paralelas de, III: 154
 «Locales», II: 77, 136
 Lógica matemática, I: 86
 Lomonosov, M. V., III: 185
 Lorentz, H. A., I: 276, 301, 304-10; III: 225
 transformación de, I: 301-3
 fórmulas de transformación de, I: 304
 Luzin, N. N., I: 83; II: 315; III: 41, 47

M

- Mach, E., III: 222
 Malsev, A. I., III: 358
 Mandelstam, L. I., I: 424; III: 254
 Máquina integradora, II: 381
 clasificadora, II: 378
 digital, II: 378
 Máquinas analógicas o de acción continua, II: 380
 Mariotte, E., I: 105, 293; II: 29; III: 174
 Markov, A. A., I: 99, 425; II: 289, 290, 306, 314, 336
 Markov, V. A., II: 314
 Markov, proceso de, II: 306
 Markusevic, A. I., III: n. 9 p. 259
 Matemática moderna, I: 87
 Matricial, álgebra, III: 389
 cálculo, III: 117
 elemento, III: 58
 notación, III: 92
 operación, III: 59

- representación, de álgebras asociativas, de grupos, III: 367
- Matriz (matrices), III: 58
 - antisimétrica, III: 397
 - canónica de Jordan, III: 108
 - de coeficientes, III: 62
 - determinante de, III: 90
 - columna de, III: 59
 - cuadrada, III: 59
 - potencias de, III: 117
 - determinante de, III: 83
 - diagonal, III: 108
- ecuación característica de, III: 106
- de igual estructura, III: 62
- fila de, III: 59
- función, III: 117
 - exponencial, III: 119
- multiplicación de, III: 63
- no degenerada, III: 94
- no nula, III: 401
- no singular, III: 94
- nula, III: 63, 64
- números característicos de, III: 106
- ortogonal, III: 397
- polinomio característico de, III: 106
- rango de, III: 95
- semejante, III: 104
- de transformación de coordenadas, III: 103
- traspuesta, III: 64
- unidad, III: 93, 391
- Máximo, I: 140
- Máximo, valor de una función, I: 141
- Maxwell, J. C., I: 21; III: 187
- Medible, conjunto, III: 43
 - función, III: 45
- Medida de un conjunto, III: 41, 43
 - de un conjunto abierto, III: 42
- Medida exterior, III: 43
- Mendeleiev, D. I., II: 336
- Menshov, D. E., II: 315
- Mergeljan, S. N., II: 340
- Meromorfa, función, II: 219
- Método, de aproximaciones sucesivas, I: 405; II: 362
 - combinatorio en topología, III: 243-244
 - de coordenadas, I: 229
 - del «corto» y «largo», I: 366
 - de las cuerdas, I: 366, 368
 - de diferencias finitas o de redes, II: 58, 60
 - geométrico, III: 197
 - de las líneas quebradas (*véase* de las poligonales)
 - de Newton, I: 368
 - numéricos, II: 355
 - de las poligonales, I: 403; II: 165, 166
 - de las potenciales, II: 54
 - de redes, II: 58, 60
 - de sustitución, I: 175
 - de las tangentes, I: 366
- Métrica, III: 201
 - intrínseca, III: 196
- Meusnier, J. B. M., II: 104, 105, 107, 108, 114, 134
 - teorema de, II: 104
- Michelson, A. A., I: 301, 302
- Miguel Angel, I: 91
- Minding, F., II: 79, 125, 137; III: 143
- Minimal(es), ideal, III: 393
 - principios, II: 130
 - problema, II: 147
 - superficie, II: 111
- Minimizadora, sucesión, II: 165
- Mínimo, I: 140
 - valor de una función, 141
 - problema del, para una integral múltiple, II: 161
- Minkowski, H., II: 139, 240; III: 223
- Möbius, A. F., I: 298; III: 237-240
 - banda de, III: 236-237
 - cinta de, III: 237
 - red de, I: 298, 299
 - superficie de, III: 236-238
- Modelo circular del plano proyectivo, III: 240-241
- Modular, retículo, III: 406
- Módulo, de un número complejo, I: 340
 - de un polinomio, I: 345
 - superficie de, I: 345, 351
- Moirve, de, II: 284
- Molin, F. E., III: 394, 395, 404

Monge, G., II: 137
 Monotonía, ley de, III: 27
 Moreland, II: 376
 Morozov, V. V., III: 358
 Morse, M., II: 138; III: 265
 Movimiento, ecuaciones de, II: 23
 helicoidal, III: 318
 impropio, III: 318, 342
 permanente, II: 183
 no perturbado, I: 418
 perturbado, I: 418
 plano-paralelo, II: 183
 de primera especie, III: 318
 propio, III: 318
 de segunda especie, III: 319
 Multiplicación, tabla de, III: 350
 Multiplicador, II: 378
 Multiplicativa, propiedad, II: 238
 Multiplicidad de una raíz, I: 353
 Mushelishvili, N. I., II: 195

N

Naimark, M. A., III: 358
 n -dimensional, ángulo, III: 170
 conjunto, III: n. 11 p. 260
 cubo, III: 176
 cuerpo, III: 181
 curva, III: 184
 espacio, III: 268-269
 espacio euclidiano, III: 77, 80
 geometría, III: 170
 plano, III: 170
 punto, III: 66
 poliedro, III: 175-181
 segmento, III: 170
 simplex, III: 178, 247
 topología, III: 259
 triangulación, III: 247
 vector, III: 66, 270
 Nash, J., II: 142
 Nasireddin Tusi, I: 61; III: 125
 Nemyski, V. V., I: 425; III: 255, 265
 Neper, J., I: 64, n. 21 p. 64, 68
 Newton, I., I: 54, 61, 62, 64, 68, 71, 73, 74, 75, 78, 96, 97, 116, 167-73, 178, 228, 237, 243, 260, 316, 317, 318, 358, 368, 369, 375,

378; II: 23, 25, 29, 78, 176, 300, 313, 320; III: 131
 método de, I: 368
 potencial de, II: 55
 teoría de los diámetros de, I: 237
 Newton-Leibniz, fórmula de, I: 169
 No enumerabilidad del continuo, III: 32
 No orientabilidad, III: 246
 Nodo(s), I: 414; III: 253
 equivalentes, III: 365
 no equivalentes, III: 365
 de una oscilación, II: 51
 Normal, II: 99
 principal, II: 95
 Novikov, P. S., III: 41, 367
 Núcleo, de una ecuación integral, III: 294
 simétrico, III: 295
 Nudos, III: 365
 equivalentes, III: 365
 no equivalentes, III: 365
 Numerabilidad, III: 27
 Número(s), I: 27
 algebraico, II: 240
 analítico, II: 340
 de Betti, bidimensional, III: 251
 característico, III: 106
 cardinal del continuo, III: 25
 compuesto, II: 242
 de dirección, I: 407
 entero, I: 24; II: 235
 negativos, I: 55
 primo, II: 235, 237
 real, axiomas de continuidad de, I: 347; III: 29, 153
 teoría analítica de, II: 340
 elemental de, II: 339
 geométrica de, II: 339
 de vueltas del campo, III: 256

O

Odner, V. T., II: 377
 Omar Khayyam, I: 61, 62
 Onda esférica divergente, II: 43
 Operación(es), sobre conjuntos, III: 19

- sobre determinantes, III: 86
- m-aria, III: 407
- no conmutativa, III: 321
- sobre vectores, III: 65
- Operador(es), III: 299
 - autoadjunto, III: 297, 302, 303
 - lineal, III: 303
 - lineal, III: 296, 299
- Orden, de conexión, III: 236, 251
 - de un conjunto de puntos, III: 33
 - de contacto, I: 404
 - de ecuaciones diferenciales, I: 384
 - de grupo, III: 325
 - de grupo simétrico, III: 327
 - de un infinitésimo, I: 151
 - de una matriz cuadrada, III: 59
 - de simetría, III: 312
 - de tangencia, I: 404
 - de transformación, III: 331
- Ordenador, I: 260
- Orientabilidad, III: 245
- Origen, I: 227
- Ortogonal(es), complementos, III: 82
 - funciones, III: 278
 - invariante, I: 289
 - matriz, III: 397
 - proyección, III: 82
 - sistema, II: 351; III: 279, 280
 - transformación, I: 287; III: 101
 - de una forma cuadrática, III: 114
- Ortogonalidad, II: 47
 - de vectores, III: 80, 270
 - propiedad de, II: 345
- Ortonormal, base, III: 80
 - sistema de funciones, III: 281
- Oscilación(es), armónica, III: 292
 - característica, II: 47; III: 293
 - ecuaciones de, II: 35
 - forzada, I: 394
 - libre, III: 292
 - modo de, II: 51
- Ostrogradski, M. V., I: 74, 99, n. 26 p. 169, 179, 203, 205; II: 58, 159
 - fórmula de, I: 203; II: 21, 163
 - Hamilton, principio de, II: 159
- P
- Papaleksi, N. D., I: 424
- Parábola, I: 248
- Parábola de orden n , I: 364
- Paraboloides, I: 273
 - elíptico, I: 269
 - hiperbólico, I: 267, 270
- Paralelas, líneas, III: 133-135
- Paralelas, postulado de, III: 124
- Paralelismo, III: 133
- Parámetro, I: 311
- Paridad, III: 328
- Parte principal, II: 220
- Parseval, M., III: 53
- Parseval, igualdad de, III: 53
- Pascal, B., II: 376
- Perforadora reproductora, II: 378
- Permutación, impar, III: 327
- Perspectiva plana, teorema fundamental de; véase Teorema de unicidad de la teoría de transformaciones proyectivas
- Peterson, K. M., II: 128, 134
 - teorema de, II: 79, 128
- Petrovski, I. G., I: 312; II: 307
- Picard, E., II: 217
- Picard, método de; véase Método de aproximaciones sucesivas
- Pitágoras, I: 40, 45, 51, 69, 229, 240; III: n. 6 p. 140, 202
- Planck, M. K. E. L., II: 307
- Plano, completado, I: 297
 - cubierto densamente, I: 299
 - ecuaciones de, I: 268
 - de las fases, I: 409
 - n -dimensional, III: 169
 - normal, II: 94
 - oscilador, II: 94
 - proyectivo, en geometría, I: 297
 - en topología, III: 239-240
 - de simetría, III: 311
 - tangente, II: 98-101
- Pogorelov, A. V., II: 141
- Poincaré, H., I: 74, 84, 99, 301, 416, 424; III: 131, 151, n. 11 p. 151, 154, 222, 243, 252, 254, 258, 260
- Poisson, S. D., II: 30-32, 43, 52, 54, 58
- Poliedros, II: 140

- en un espacio n -dimensional, III: 175
 Polinomio(s), característico, de una matriz, III: 106
 gráfica de, I: 364
 de interpolación, II: 316, 335
 irreducibilidad de, I: 334
 de Legendre, II: 330; III: 282
 de una matriz, III: 117
 simétrico, I: 328; III: 320
 simétrico elemental, I: 328
 trigonométrico, II: 314
 Polo II: 219
 Poncelet, J. V., III: 160
 Pontriagin, L. S., I: 424; III: 260.
 n. 12 p. 260, 261, 311, 359
 ley de dualidad de, III: 261
 Popov, A. S., I: 21
 Posición falsa, método de, I: 366, 368
 Postnikov, M. M., III: 261
 Potencial, de capa simple, II: 56
 de fuerzas, II: 55
 de velocidad, II: 27, 187
 de un fluido, II: 26
 Primitiva, función, I: 168
 Primo, gemelo, II: 242
 Principio(s), del argumento, I: 360
 de Arquímedes, III: 31-32
 de continuidad de Cantor, III: 30
 de Dedekind, III: 29
 Probabilidad, II: 269, 273
 adición de, II: 273
 densidad de, II: 306
 distribución de, II: 286
 Proceso aleatorio, II: 299, 306
 Proclo, III: 125
 Producto, III: 26
 de caminos, III: 361
 de ciclos, III: 328
 derivada de, I: 133
 libre, III: 364
 de matrices, III: 60
 no conmutativa, III: 60
 de transformaciones, III: 321
 Prolongación analítica de una función en un dominio complejo, II: 170-171, 222-225
 Propagación de ondas, II: 40
 Propiedad intrínseca, II: 127
 Proporcionalidad, III: 56
 Proximidad, III: 359
 Proyección perspectiva, I: 293-5
 Proyectiva(s), aplicaciones: teorema fundamental, I: 298
 transformación, I: 298; II: 143
 Punto(s)
 adherente, III: 262-264
 aislado, III: 35
 conjunto de, III: 32
 conjunto lineal de, III: 32
 coordenadas de, I: 227
 de discontinuidad, I: 411
 de ensilladura, I: 413
 de estancamiento, II: 193
 fijo, III: 257
 frontera, III: 35
 impropio, III: 239
 interior de un conjunto, III: 35
 de interpolación, II: 320
 de intersección, I: 230
 de inflexión, I: 148
 límite, III: 35
 ordinario, I: 410
 plano, II: 103
 simétrico, III: 311
 singular, III: 253, 256
 aislado, I: 410
 de tangencia, II: 99

R

- Racionales, números, III: 26
 Radical, III: 394
 de un anillo, III: 404
 Radio de convergencia, I: 220; II: 173; III: 160
 Radio vector, II: 130
 Rafael, I: 91
 Raíces, cálculo de, I: 364
 múltiples, I: 353
 multiplicidad de, I: 353
 reales, I: 327
 Rama uniforme, II: 182, 227
 Ramanujan, S., II: 240
 Ramificación, punto de, de función compleja, II: 182, 226
 Rango, de álgebra, III: 387
 de una matriz, III: 95

- de números hipercomplejos, III: 376
 - de un sistema de vectores, II: 75
 - Rasevski, P. K., III: 215
 - Real, espacio, III: 215
 - número, I: 47; III: 25, 28
 - parte de un cuaternio, III: 381, 382-383
 - punto, I: 298
 - raíz, I: 327
 - recta, I: 298
 - Reciprocidad, ley de, III: 289
 - Recta, línea, II: 123
 - Recta numérica, III: 16, 29
 - Redes de curvas, II: 79
 - Reducción, a una forma canónica, I: 253
 - de una forma cuadrática, III: 111
 - representación, III: 370
 - Reducible, ecuación, I: 292
 - Regular, representación, III: 371
 - sistema de puntos espaciales, III: 342
 - superficies, II: 97
 - Representación completamente reducible, III: 370
 - Representación fiel, III: 391
 - Representación(es), de álgebras asociativas, III: 391
 - equivalentes, III: 367
 - de grado n , III: 368
 - de grupo, III: 367
 - irreducible, III: 370
 - isomorfa, III: 391
 - Reproductora, II: 378
 - Resolubilidad, de grupo, III: 334
 - de problemas con regla y compás, I: 335
 - mediante radicales, III: 346
 - Resoluble, grupo, III: 334
 - mediante radicales, I: 337
 - Retículo, III: 404
 - modular, III: 406
 - «Resto», término, I: 159
 - Riccati, J. F., I: 395
 - ecuación de, I: 395
 - Riemann, B., I: 80, 81, 99, 121, 312; II: 180, 183, 187, 194, 198, 201-207, 218, 227-230, 240, 246; III: 16, 45, 47, 49, n. 7 p. 144, 197-214, 216, 221, 222, 226, 258, 357
 - espacio de, III: 199
 - geometría de, III: 199
 - hipótesis de, II: 246
 - integral de, I: 165; III: 47
 - métrica de, III: 205
 - superficie de, II: 201-207; III: 258
 - Riesz, F., I: 84
 - Ritz, W., II: 73-74, 359
 - Rolle, M., I: 354, 356
 - teorema de, I: 354
 - Rotación, del espacio, III: 318
 - de una transformación, II: 197
 - Rotaciones, grupo de, III: 326
 - Rotacional, flujo, II: 26
- S
- Saccheri, G. G., III: 126, n. 2 p. 126, 128, 131
 - Safarevich, I. R., I: 338; III: 346
 - Schauder, J., III: 265
 - Schlaefli, L., III: 176
 - Schrödinger, E., II: 73
 - Schwartz, desigualdad de; véase Desigualdad de Cauchy-Bunjakovski
 - Schweikart, III: 127, 131
 - Sección normal, II: 102
 - Segmento parabólico, I: 404
 - Selberg, A., II: 246
 - Semiintervalo, III: 33, 39
 - Semiinvariante, I: 290
 - Semirretículo, III: 405
 - inferior, III: 405
 - superior, III: 405
 - Separación de variables, II: 44
 - Series, I: 206
 - absolutamente convergente, I: 211
 - armónicas, I: 210
 - divergentes, I: 207
 - no uniformemente convergentes, I: 213
 - de potencias, I: 217; II: 169
 - sumables, I: 207, n. 37 p. 207
 - de Taylor, I: 162
 - trigonométricas, III: 283

- uniformemente convergentes, I: 212
- Serre, J. P., III: 261
- Serret, J. A., I: 318
- Seudoesfera, III: 142
- Símbolos, I: 27
- Simetría, III: 311, 317
- bilateral, III: 312
- central, III: 312
- centro de, III: 312
- circular dirigida, III: 336
- eje de, III: 312
- especular, III: 311
- de orden n , III: 312
- respecto a ejes, III: 342
- Simplex, III: 178, n. 22 p. 178
- orientado, III: 250
- Simpson, T., II: 324-330, 360, 368, 369
- fórmula de, II: 325
- Singular, punto, I: 410; III: 256
- Sistema físico, oscilaciones características, I: 391; III: 293
- Sistema, algebraico, III: 309, 406-408
- binario, III: 404
- de ecuaciones lineales, III: 83
- homogéneo de ecuaciones lineales, III: 96; véase también
- Ecuaciones de primer grado
- de intervalos, III: 29
- normal, II: 351
- ortogonal de funciones, III: 282, 285
- normal, II: 351
- posicional, I: 29
- Sitnikov, K. A., n. 12 p. 260, 261
- Smidt, O. Ju., III: 311, 373
- Smirnov, Ju. M., III: 266
- Smirnov, V. I., I: n. 33 p. 194
- Snirelman, L. G., II: 138; III: 260
- Sobolev, S. L., II: 69, 73
- Solución(es), admisible, II: 68, 69
- algebraica, I: 320
- de ecuaciones diferenciales, I: 384
- generalizada, II: 68, 69
- unicidad de, I: 401; II: 29
- Stepanov, V. V., I: 425; II: 54, 302; III: 255
- Stevin, S., I: 64
- Sturm, S., I: 318, 358
- teorema de, 357
- Subálgebra, III: 393
- Subanillo, III: 399
- Subespacio, III: 74
- Subespacio generado por un sistema de vectores, III: 74-75
- Subgrupo(s), III: 330-334
- generadores de, III: 330
- invariante, III: 352
- normal, III: 351
- normales no triviales, III: 358
- simple, III: 358
- Subnikov, A. V., III: 343
- Sucesión, convergente en casi todo punto, III: 46
- fundamental, III: 30
- minimizante, II: 16
- de números reales, III: 30-31
- Suceso(s), aleatorio, II: 271, 296, 303
- cierto, II: 269
- estocástico, II: 271
- imposible, II: 269
- independiente, II: 275
- intersección de, II: 274
- mutuamente excluyentes, II: 272
- necesario, II: 269
- unión, II: 272
- Suma, de conjuntos, III: 19
- derivada de una, I: 132
- de Fejér, II: 349
- parcial, I: 207
- de series, I: 207
- trigonométrica, II: 247, 248
- de vectores, I: 262
- Sup., I: 213, n. 39 p. 213
- Superficie(s),
- analítica, II: 125
- cerrada(s), III: 237, 245, 246
- clase de, II: 139
- convexa, II: 140-141
- orientable, III: 250-251
- con curvatura variable de Gauss, II: 121
- conexión de, III: 236
- congruente, II: 128
- de curvatura media nula, II: 111
- de dos caras, III: 236
- de una sola cara, III: 236
- deformabilidad de, II: 142

- deformación de, II: 115, 125, 142
 - desarrollable, II: 120
 - de discontinuidad, II: 68
 - ecuaciones de, I: 266
 - familia de, II: 134
 - género de, II: 230; III: 236
 - geodésica, III: 210
 - lisa, II: 99
 - mínima, II: 111
 - de Möbius, III: 236-237
 - no cerrada convexa, II: 142
 - no lisa, II: 140
 - no regular, II: 137
 - normal, III: 236
 - regular, II: 97
 - de revolución de menor área, II: 147
 - teoría de, II: 96
 - Superposición, método de, II: 44
 - Suslin, M. Ja., III: 41
 - Sustracción, III: 26
 - Sylvester, J. J., I: 358; II: 245
- T
- Tablas, II: 273
 - de dos entradas, II: 374
 - de una entrada, II: 374
 - Tabulador, II: 378
 - Tales, I: 40
 - Tangente, I: 122; II: 86, 87
 - Tangentes, campo de, I: 399
 - Tartaglia, N., I: 63, 318, 321, 322
 - Taurinus, III: 127, 131
 - Taylor, B., II: 320, 321
 - fórmula de, I: 157, 191
 - serie de, I: 162
 - Temperatura, gradiente de, II: 28
 - Tensor, II: 16
 - Teorema,
 - de Abel, I: 332; II: 171
 - de Bolzano-Wierstrass, I: 347; III: 36
 - de Bonnet, II: 128
 - de Budan, I: 357
 - de Cauchy, II: 210
 - central del límite, II: 290
 - de Codazzi, II: 80, 128
 - de contracción, II: 364
 - de dualidad, III: 260
 - de Euler, II: 103, 105
 - fundamental, del álgebra, I: 338-352; II: 169
 - de la perspectiva plana, I: 295-297
 - de Gauss; véase Ostogradski, fórmula de
 - de Green, II: 163
 - de Hopf, III: 256
 - del límite central, II: 290
 - de Meusnier, II: 104
 - de Peterson, II: 79, 128
 - de Picard, II: 217
 - del punto fijo, de Brouwer, III: 198, 258
 - de Rolle, I: 354
 - de Sturm, I: 357
 - de Sukovski, II: 190, 193
 - de unicidad, para funciones analíticas, II: 222
 - de la teoría de transformaciones proyectivas, I: 294
 - del valor medio, I: 154
 - de Wedderburn, III: 395
 - de Weirstrass, II: 337
 - de Wilson, II: 265
 - Teoría, algebraica de números, II: 240
 - analítica de números, II: 240
 - cualitativa de ecuaciones diferenciales, I: 415; III: 253-254
 - de curvas, II: 80
 - elemental de números, II: 240
 - de funciones de variable real, III: 16-17
 - de Galois, I: 332
 - geométrica, de funciones de números, II: 240
 - de Newton, de diámetros, I: 237
 - de números, II: 235
 - de operadores, III: 297
 - Tihonov, A. N., III: 266
 - Tolomeo, C., I: 58, 60, 63
 - Topológico(a), aplicación, III: 263
 - espacio, III: 193, 262, 265
 - estructura, III: 261
 - grupo, III: 358
 - imagen, III: n. 10 p. 259
 - invariante, III: 247, 259, 362
 - producto, III: 242

- propiedad, III: 232
 - transformación, III: 232
 - Topología, III: 231
 - ciclos en, III: 249
 - combinatoria, III: 260
 - método de, III: 243-244
 - conjuntista, III: 259
 - n -dimensional, III: 259
 - orientaciones compatibles en, III: 245
 - de superficies cerradas, III: 258
 - Toro, III: 241
 - de una cara, III: 238
 - Torricelli, I: 97
 - Torsión, II: 95; III: 250
 - Tractriz, III: 142
 - Trascendente, función, II: 217
 - número, II: 240
 - Transformación(es), afín, I: 280; III: 162
 - bicontinua, III: 362
 - conforme, II: 197, 200
 - conjugada, III: 323, 324, 332, 333
 - de coordenadas, I: 255; III: 102
 - cuasi-conformes, II: 204
 - degenerada, III: 101
 - general afín, I: 280
 - grupo de, III: 351
 - de grupo, III: 163
 - grupos continuos de, III: 336
 - homogénea, III: 99
 - identidad, III: 330
 - inversa, III: 99
 - lineal, III: 98
 - de Lorentz, I: 301, 303
 - matriz de los coeficientes de, III: 100
 - no degenerada, III: 101
 - orden de, III: 331
 - ortogonal, I: 287; III: 101
 - periódica, III: 331
 - producto de, III: 321
 - proyectiva, I: 298; II: 143; III: 158
 - simétrica, III: 314
 - singular, III: 101
 - no singular, III: 101
 - topológica, III: 232
 - unitaria, III: 358
 - uno a uno, III: 315
 - Transposición, III: 328
 - Traslación, III: 318
 - Traspuesta, III: 64
 - matriz, III: 102
 - Traza, III: 372
 - Treffitz, J., II: 73
 - Triangulación, III: 244
 - tridimensional, III: 247
 - Triangular, desigualdad, III: 196
 - ley, III: 264
 - Tschirnhausen, E. W., I: 318, 326
- U
- Ulug Begh, I: 62
 - Unicidad, de descomposición, III: 401
 - propiedades de, II: 222
 - de solución, I: 401; II: 29
 - Unicidad, teorema de, para funciones analíticas, II: 222
 - de la teoría de transformaciones proyectivas, I: 295
 - Unidad, III: 401
 - Uryson, P. S., III: 259, 260, n. 11 p. 260, 263, 266
- V
- Vagner, V. V., III: 215
 - Vallée-Poussin, C. J., de la, II: 314
 - Van der Waerden, B. L., I: 320
 - Variable, I: 65
 - aleatoria, II: 286
 - cambio de, I: 175
 - continua, I: 54
 - dependiente, I: 103
 - independiente, I: 103
 - separación de, II: 44
 - Variable, magnitud, I: 65, 110
 - Variación, II: 154
 - Variacional, problema, II: 157
 - Varianza, II: 287
 - Varichak, III: 151
 - Variedades, III: 241-243
 - aplicación continua de, III: 257

- conexión de, III: 251
 - irreducible, III: 403
 - lisas, III: 255
 - multi-dimensionales, III: 259
 - Vavilov, S. I., I: n. 1 p. 92
 - Veblen, O., III: 260
 - Vector(es), I: 261; III: 65-66
 - combinación lineal de, III: 70
 - de un espacio infinito-dimensio-
nal, III: 275
 - flujo, II: 20
 - inverso de un, I: 262
 - longitud de, III: 80
 - n -dimensional, III: 270
 - dos a dos, III: 272
 - velocidad, II: 184
 - Vectorial(es), álgebra, I: 261; III: 381
 - campo, II: 15; III: 252
 - continuo, III: 252
 - coordenadas, I: 263; III: 73
 - función, II: 130
 - operaciones, III: 65
 - parte de un cuaterno, III: 380, 382
 - suma, I: 262
 - de dos subespacios, III: 75
 - Velocidad potencial, II: 27, 187
 - Vieta, F., I: 61, 63, 317, 319, 326, 328; III: 344
 - Vinogradov, I. M., I: 87; II: 240, 241, 246, 256-264, 316; III: 402
 - Vonoroï, G. F., II: 241
- W
- Wachter, III: 127
 - Wallace, A. R., III: 131
 - Wallis, J., III: 125
 - Wedderburn, III: 394, 395, 404
 - teorema de, III: 395
 - Weierstrass, K. T., I: n. 14 p. 46, 83, 99, 129; II: 314, 337; III: 30, 31, 36
 - principio de, III: 30
 - teorema de, II: 337; III: 36
 - Weyl, H., III: 398
 - Wilczynski, E. J., II: 143
 - Wilson, teorema de, II: 265
- Z
- Zenón, I: 46
 - Zolotarev, E. I., I: 318; II: 240, 314; III: 402
 - Zukovski, N. E., I: 344; II: 190, 191, 193
 - teorema de, II: 190, 193